

J. Kutscher

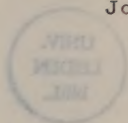
München 1976

Untersuchung der
empiristischen Rekonstruktion
von Urteilen über Schüler

Inaugural - Dissertation
zur Erlangung des Doktorgrades
der Philosophie
der Ludwig-Maximilians-Universität zu München

vorgelegt von

Joachim Kutscher in München



München, Eigendruck

1977

Referent: Prof. Dr. Hans Schiefele

Korreferent: Prof. Dr. Kurt Lukascyk

Tag der mündlichen Prüfung: 26.02.1976



Inhaltsverzeichnis

Vorwort	6
1. Die Funktion der "Objektivität" in der Schülerbeurteilung	11
1.1 Entwicklungstendenzen in der Verwissenschaftlichung der Schülerbeurteilung	11
1.2 Die Objektivitätsforderung als Konsequenz unbewältigter Probleme	17
1.2.1 "Objektivität" als formalisierte Inter-subjektivität	23
1.2.2 "Objektivität als Übereinstimmung im Gebrauch einer Beobachtungssprache	29
1.2.3 "Objektivität" als intersubjektiv konsistente Deutung von invarianten Verhaltensmustern	39
1.3 "Objektivität" in der Diagnostik: Ausdruck der Entsubjektivierung des Forschungsprozesses ohne Gegenstandsanalyse	49
1.3.1 Zwei exemplarische Funktionen des Entsubjektivierungsparadigmas im Forschungsprozeß	49
1.3.2 Historische Bedingungen für das Wirksamwerden des Entsubjektivierungsparadigmas im Bereich der Diagnostik in der Schule	57
1.3.2.1 Die psychotechnisch-diagnostische "Bewährungsprobe" des Entsubjektivierungsparadigmas in außerschulischen Bereichen	62

1.3.2.2	Die Objektivierung der Kritik am Lehrerurteil durch das Entsubjektivierungsparadigma	65
1.3.2.3	Vorläufige Zusammenfassung: Entsubjektivierung als formales Prinzip	68
1.3.2.4	Das Forschungsparadigma der Entsubjektivierung als Ausdruck der Technologisierung gesellschaftlicher Prozesse . .	73
1.4	Der Stellenwert von Forschungsparadigmen im Wissenschaftsprozess	85
1.4.1	"Irrationalität" in der Wissenschaft ?	86
1.4.2	Darstellung des KUHNschen Paradigmagedankens	88
1.4.3	Kritische Einwände gegenüber KUHN und ihre Bedeutung für die Entsubjektivierung der Diagnostik	101
1.5	Zusammenfassung und Problemstellung . .	114
2.	Der Gegenstandsbereich der Schülerbeurteilung	119
2.1	Empirische Bedeutung und empiristische Rekonstruktion von Dispositionen . . .	122
2.2	Orientierungsrahmen für die Rekonstruktion dispositioneller Schülereigenschaften: Allgemeine Bestimmung des Anwendungsbereiches pädagogischer Diagnostik	135
2.2.1	Zwei Beispiel für die Rekonstruktion dispositioneller Eigenschaften	139
2.2.2	Formale Bestimmung des Anwendungsbereiches	150
2.3	Verfahrenstechnische Festlegung von Dispositionen durch die Zensurengebung	156
2.3.1	Das Dilemma in der Interpretation von Beurteilungskategorien	158

2.3.2	Formale Synthese dispositioneller Eigenschaften: die Funktionalisierung des Anwendungsbereiches pädagogischer Diagnostik	168
2.4	Der Verlust des Anwendungsbereiches pädagogischer Diagnostik und die gemeinsame Verständigungsbasis verschiedener Beurteiler	189
2.4.1	Die kompliziertere Struktur diagnostischer Dispositionsprädikate	191
2.4.2	Die fehlende Verfügung über eine Beobachtungssprache	194
2.4.3	Zusammenfassung	201
2.4.4	Die Festlegung des Beobachtbaren	206
2.4.5	Die Standardisierung der Schülerbeurteilung als Ersatz für ihre empirische Begründung	224
2.4.5.1	Die "Vorzüge" der objektivierten Schülerbeurteilung gegenüber der Zensurengebung	227
2.4.5.2	Meßinstrumentelle Bedingungen objektivierter Schülerbeurteilung	234
2.4.5.3	Das Problem der empirischen Deutung von "Meßwerten" in der Schülerbeurteilung	239
2.4.5.4	Die Entsubjektivierung des Deutungsproblems	253
2.4.5.4.1	Verfahrenstechnische Festlegung der Itemäquivalenz	255
2.4.5.4.2	Die Formalisierung der Deutung von Schülerbeurteilungen	267
3.	Schlußfolgerungen	280
3.1	Ergebnisse der Versuche zur Entsubjektivierung der Schülerbeurteilung	280

3.2	Zur realen Notwendigkeit der Entsubjektivierung der Schülerbeurteilung . . .	294
-----	--	-----

	Literaturverzeichnis	299
--	--------------------------------	-----

	Lebenslauf	309
--	----------------------	-----

Vorwort

Die Idee zur vorliegenden Arbeit entstand im Verlauf meiner Tätigkeit im Forschungsprojekt "Schullaufbahnberatung" am Institut für Pädagogik II der Universität München und am Institut für Unterrichtsforschung der Pädagogischen Hochschule Augsburg. Die ursprüngliche Aufgabe des Forschungsprojektes bestand in der Entwicklung und Erprobung von standardisierten Beurteilungsverfahren für die Gewinnung von diagnostischen Informationen zur wissenschaftlichen Begründung von Übertrittsentscheidungen. Für die an einigen Schulen in Bayern eingeführte Orientierungsstufe sollte Beurteilungsinstrumente geschaffen werden, die zu einer "objektiven" Beurteilung des Lernverlaufes und Lernerfolges der Schüler in der Orientierungsstufe beitragen und die Verteilung von Schülern auf die zur Verfügung stehenden Selektionsalternativen "Hauptschule", "Realschule" und "Gymnasium" wenigstens partiell auf eine wissenschaftliche Grundlage stellen.

Der Versuch, die Schülerbeurteilung durch psychodiagnostische Methoden zu optimieren und den Beurteilungsprozeß empirisch zu fundieren, führte zu keinen, über den damaligen Stand psychologisch-pädagogischer Forschung hinausreichenden Verbesserungen selektionsorientierter Schülerbeurteilung. Auch die gegenwärtig vorliegende, wissenschaftliche Literatur zu diesem Fragenkomplex verzeichnet keinerlei Fortschritt in der Beantwortung der Frage nach einer Objektivierung der praktizierten Diagnostik in der Schule, die der Forderung nach zuverlässigen und im gewünschten Maß prognostisch gültigen Urteilen gerecht wird. Dies ist umso bemerkenswerter, als Wissenschaftler in der Bundesrepublik seit über zehn Jahren damit beschäftigt sind, das als subjektivistisch kritisierte Lehrerurteil traditioneller Art durch sog. objektive Beurteilungsverfahren zu ersetzen. Überdies erbrachten auch die entsprechenden, wesentlich älteren Unternehmungen in angloamerikanischen Ländern nicht die gewünschten Erfolge. Nachwievor werden Urteile über Schüler in mehr oder weniger modifizierten Zensurenformen ausgedrückt.

Für die Beibehaltung der Zensurengebung zum Zweck der Verteilung der Schüler auf weiterführende Ausbildungsalternativen sind eine Reihe von komplexen Ursachen verantwortlich zu machen, deren historisch-systematische Untersuchung noch ansteht. Nicht zuletzt spielt dabei das Scheitern in der Einlösung des Anspruchs eine Rolle, durch die methodische Objektivierung des diagnostischen Prozesses mittels standardisierter Beurteilungsverfahren zu eindeutig interpretierbaren und intersubjektiv konsistent gebrauchten Urteilen über Schüler zu gelangen. Der Anspruch wird in der wissenschaftlichen Literatur empiristischer Provenienz trotz aller feststellbaren Mißerfolge mit bisweilen dogmatischem Anstrich als erfolgsverheißendes Forschungsvorgehen ungebrochen vertreten. Diese Arbeit versucht einen Beitrag zur Beantwortung der Frage zu leisten, ob der Anspruch allgemein gerechtfertigt ist und ob sich der eingeschlagene Weg der methodischen Objektivierung und Standardisierung der Beurteilertätigkeit für die Optimierung der pädagogischen Diagnostik als erfolgsträchtig erweisen kann.

Die Apologeten der Verwissenschaftlichung der Schulleistungsbewertung beziehen sich in ihren Argumentationen auf die in vielen Untersuchungen nachgewiesene, mangelnde Objektivität der traditionellen Zensurengebung und auf die damit verbundene Ungerechtigkeit einer rational nicht begründbaren Vergabe von unterschiedlichen Bildungs- und Lebenschancen an die betroffenen Schüler. Diese Kritik an den Zensuren ist sicherlich berechtigt. Sie verliert aber ihre Funktion als Aufforderung zu einer wissenschaftlich begründbaren Neugestaltung praktizierter Diagnostik in der Schule, wenn die Objektivierungsversuche in der Schulleistungsbewertung Ergebnisse erbringen, die der Objektivierungsforderung nur der Form nach entsprechen und die Entscheidung über die Vergabe von Bildungschancen lediglich instrumentalisieren. Der gegenwärtige Forschungsstand auf dem Gebiet der pädagogischen Diagnostik provoziert den Verdacht, daß sich die Versuche zur Verwissenschaftlichung schulischer Diagnostik in einem derartigen Dilemma befinden.

Die Objektivierungsbestrebungen im Bereich der Schülerbeurteilung

lassen sich durch die Orientierung der Diagnostikergemeinschaft an zwei wissenschaftsmethodischen Normen kennzeichnen: Urteile über Schüle müssen intersubjektiv konsistent interpretiert werden können, d.h. "objektiv" sein, und ihre inhaltliche Bedeutung muß durch wissenschaftlich saubere Untersuchungen "empirisch abgesichert" sein. In der Orientierung an beiden Normen erscheint das Bekenntnis empirisch-sozialwissenschaftlicher Forscher im Bereich der Diagnostik in der Schule zu einer empiristischen Form von Wissenschaftstheorie. In der Entwicklung des "Modernen Empirismus" und in der Diskussion der Theorienentwicklung und -begründung durch die "Analytische Philosophie" stehen die Fragen des intersubjektiven Gebrauchs von wissenschaftlichen Aussagen und ihres empirischen Gehaltes im Mittelpunkt. Ihre Behandlung und Problematisierung basiert auf der Reflexion des exakt-naturwissenschaftlichen Forschungsprozesses. Von daher betrachtet wäre der grundsätzliche Einwand zu untersuchen, ob die erarbeiteten Formen des empiristisch-wissenschaftlichen Erkenntnisgewinns auf Fragestellungen des Sozialwissenschaften übertragen werden können. Die präzise und detaillierte Analyse des Einwandes setzt aber Vorarbeiten voraus, die sich beim gegenwärtigen Stand der Erörterung wissenschaftlicher Erkenntnisproblematik bestenfalls im Anfangsstadium befinden oder mit der Intention, die Unverträglichkeit der Formen naturwissenschaftlicher und sozialwissenschaftlicher Erkenntnis nachzuweisen, pauschalierend übergangen werden. In solchen Vorarbeiten wäre die historische Rekonstruktion der Beziehung zwischen Forschungsgegenstand und Forschungsmethode in den verschiedenen Natur- und Sozialwissenschaften zu leisten und der Zusammenhang zu den jeweils geltenden Erkenntnisprinzipien und -paradigmen herzustellen.

Für die folgende Betrachtung der praktizierten Diagnostik in der Schule und die Untersuchung der Ursachen für die relative Erfolglosigkeit der Objektivierungsbemühungen soll deshalb von den bestehenden Gewohnheiten in der Verwissenschaftlichung der Schülerbeurteilung ausgegangen werden. Unter Berufung auf verschiedenste Reflexionsansätze empiristischer Wissenschaftstheorie legitimieren Diagnostiker die Entwicklung und Anwendung von

standardisierten Verfahren zur Schülerbeurteilung durch bestimmte, subjektunabhängige Konstruktions- und Auswertungsmethoden. Das Problem der inhaltlich-empirischen Voraussetzungen für die Anwendung der Methoden und ihrer Adäquanz für eine prognostisch gültige Beurteilung wird in der Regel vernachlässigt oder nicht gesehen. In der vorgelegten Arbeit wird die Absicht verfolgt, die Vertreter der Objektivierungsbestrebungen im Rahmen der Schülerbeurteilung beim Wort zu nehmen und ihr konkretes Vorgehen an ihrem eigenen Anspruch empiristischer Wissenschaftlichkeit zu bemessen.

Da die Untersuchung in erster Linie Urteile über Schüler in Form von diagnostischen Aussagen über Qualifikationen zum Gegenstand hat, wird sich der Vergleich des konkreten Forschungsvorgehens mit dem zugrundegelegten Anspruch nur auf ausgewählte Aspekte der empiristischen Wissenschaftstheorie beziehen können. Zudem ermöglicht die Tatsache, daß Urteile über Schüler im Kontext des bestehenden Schulsystems zunächst und vorwiegend im Dienst der Selektionsfunktion stehen, eine Spezifizierung der Betrachtungsperspektive. Der Schwerpunkt der Analyse liegt deshalb auf der Diskussion der Frage, über was sich die Diagnostikergemeinschaft unterhält, wenn sie von "objektiven" Urteilen spricht, und inwiefern solche Urteile empirisch bedeutungshaltige Aussagen über dispositionelle Eigenschaften von Schülern sein können. Da die objektivierte Beurteilung den subjektivistischen Urteilsprozeß des Lehrers ablösen soll, ist zugleich darzustellen, auf welche Bedingungen der unterstellte Objektivitätszuwachs beim Übergang von Zensuren auf objektivierte Urteilsdaten zurückgeführt werden kann.

Auf einen zusammenfassenden Nenner gebracht, setzt sich die Studie mit dem Problem der prognostischen Relevanz von angeblich objektiven Schülerbeurteilungen und der Möglichkeit auseinander, die These, die Schülerbeurteilung vom Subjektivismus des urteilenden Lehrers zu befreien, empiristisch wissenschaftlich zu begründen. Es wird sich dabei zeigen, daß die These nur um den Preis einer naiven Verkürzung empiristischer Vorgehensweise in der Verwissenschaftlichung des Urteilsprozesses zu halten ist.

Meine Motive für die Beschäftigung mit der wissenschaftlichen Begründbarkeit selektionsbezogener Schülerbeurteilung entwickelten sich auf dem Boden der Erfahrungen, die ich im Rahmen der konkreten Versuche sammeln konnte, im Forschungsprojekt "Schul-
laufbahnberatung" an der Konzeption von objektivierten Beurteilungsverfahren zu arbeiten. Die unmittelbare Konfrontation mit der Diskrepanz zwischen den Ergebnissen der Anwendung einer erfolgsverheißenden, wissenschaftlichen Methodik und ihrer unreflektierten, erwarteten Brauchbarkeit für die Optimierung pädagogischer Diagnostik forderten zu einer grundsätzlichen Betrachtung des Verhältnisses von Anspruch und Effektivität heraus. Die Auseinandersetzung mit den verschiedenen Bedingungen für die feststellbaren Schwierigkeiten bei der Schülerbeurteilung stellt selbst auf der von mir gewählten Ebene nur eine partielle Problembehandlung dar.

Während der Arbeit an diesem Thema erhielt ich von verschiedenen Seiten ein Fülle von Anregungen, die mich in der konsequenten Fortsetzung des eingeschlagenen Weges bestärkten. Ich möchte mich deshalb bei all jenen bedanken, die sich für meine Ideen interessiert haben. Mein besonderer Dank gilt meinem Doktorvater, Herrn Prof. Dr. Hans Schiefele, der mich durch seine intensive Beschäftigung mit meinem Vorhaben immer wieder ermutigte und mir die Gelegenheit gab, in mehreren Seminaren die Sinnhaftigkeit meiner Vorstellungen durch Diskussionen zu überprüfen. Von entscheidender Bedeutung für die Initiative zu dieser Untersuchung waren schließlich die konsequenten Streitgespräche im Diskussionszirkel "Wienerwaldkreis" über Grundprobleme empirisch pädagogischer Forschung.

1. Die Funktion der "Objektivität" in der Schülerbeurteilung

1.1 Entwicklungstendenzen in der Verwissenschaftlichung der Schülerbeurteilung

Der Problemkreis "Pädagogische Diagnostik" hatte in den letzten Jahren im Rahmen der Bemühungen um eine Reform des Bildungswesens eine zunehmende Akzentuierung erfahren. Im Bildungsgesamtplan, den die Bund-Länder-Kommission für Bildungsplanung am 15.6.1973 verabschiedet hat, wird neben der Curriculumforschung, der Bildungstechnologie und Bildungsökonomie, die Pädagogische Diagnostik als wesentlicher Problembereich hervorgehoben. Gleichzeitig haben die Ergebnisse der Curriculumforschung den inneren Zusammenhang von Curriculumevaluation und pädagogischer Diagnostik verdeutlicht. Pädagogische Diagnostik als umfassender Begriff für den Zusammenhang von Schülerbeurteilung und Unterrichtsbeurteilung ist in der Zwischenzeit zu einer eigenständigen Kategorie in der pädagogisch-psychologischen Forschung geworden. Die Erforschung der pädagogischen Diagnostik wird auch in den Zielvorstellungen der Bundesregierung zur Curriculumentwicklung im Bildungsbericht '70 als Aufgabe genannt. Dort heißt es unter dem Titel "Forschung und Entwicklung": "Eine systematisch angelegte und effektive Lehrplanreform und Curriculumentwicklung setzt den Aufbau eines verbundenen Systems der Curriculumforschung, -entwicklung und -erprobung voraus. Grundlagenforschung, Entwicklung und Auswertung sind mit der Testforschung und -entwicklung sowie mit der Entwicklung und Erprobung von Unterrichtstechnologien zu koordinieren" (S. 135).

Die in der Zwischenzeit wieder partiell zurückgenommene, forschungsprogrammatische Profilierung der pädagogischen Diagnostik ist Ausdruck eines offiziösen Interesses an der Verwissenschaftlichung des Bildungswesens und Reaktion auf die defizitäre Forschungslage auf dem Gebiet der praktizierten Diagnostik in der Schule. Trotz der gesellschaftspolitischen Bedeutung des Problems kann die pädagogisch-psychologische Forschung bisher keine

systematisch Reflexion der Schwierigkeiten der praktizierten Diagnostik in der Schule vorweisen und ist infolgedessen weder zu einem ansatzweise theoretisch begründeten noch zu einem pragmatisch relevanten Modellentwurf vorgedrungen. "Selbst wenn man die gelegentlich masochistisch wirkende Neigung zur Selbstkritik mancher Diagnostiker in Rechnung stellt, wird man dem Urteil zustimmen müssen, daß die Forschung, gemessen an den theoretischen und praktischen Bedürfnissen, noch ganz in den Anfängen steckt... Zudem gewinnt man den Eindruck, daß die berichteten Forschungsarbeiten ganz unsystematisch und unkoordiniert die Interessen der einzelnen Forscher widerspiegeln. Auch sind die bisherigen Versuche, neue Modell der pädagogischen Diagnostik zu entwickeln und zu erproben, vereinzelt geblieben oder stecken noch in den Anfängen..." (HOPF, 1975, S. 251). Für dieses Dilemma sind nach meiner Auffassung thesenartig folgende Begründungen anführbar:

1. Die Forschung im Bereich der Diagnostik in der Schule war und ist in erster Linie auf die Selektionsfunktion ausgerichtet und wirkt dadurch der Entwicklung ihrer eigentlichen inhaltlichen Grundlage, der Curriculum- und Unterrichtsforschung entgegen.
2. Durch die damit verbundene Trennung und Verselbständigung von Unterricht und Diagnostik konnte die Forschung nicht mehr auf das Erfahrungsfeld "Schule" rekurrieren und bezog ihre Impulse aus anderen Forschungsfeldern (Werbung, Militär, Betriebswirtschaft), deren Problemstellungen lediglich formal auf den Selektionsaspekt im Bildungswesen zu übertragen waren (vgl. HALLER, 1973).
3. Dies ermöglichte in der Folge die Konzentration der Forschungsaktivität auf die methodische Optimierung der Schülerbeurteilung und mußte in dem permanenten Versuch münden, den inhaltlichen Rückbezug der Beurteilungsmethoden nachträglich zu legitimieren.
4. Zur Legitimation bezog man sich auf wissenschaftliche Begründungsmuster der empiristischen Wissenschaftstheorie, ohne die dort ausgewiesenen Anwendungs- und Gültigkeitsbedingungen zur Kenntnis zu nehmen, deren Berücksichtigung die Voraussetzung für den Anspruch auf die Exaktheit empiristisch wissenschaftlichen Arbeitens darstellt.

In der wissenschaftlichen Literatur hat man sich vorwiegend um eine Kritik an der Selektionsfunktion diagnostischer Verfahren und an den methodischen Optimierungsansätzen bemüht. Ich glaube

jedoch, daß die Behandlung dieser Fragen zum Teil zu pauschalierend erfolgte, zum Teil sich zu sehr an isolierten Details vollzog, um problemlösende und praxisrelevante Forschungsprozesse zu initiieren.

Im Gegensatz dazu verfolgt diese Arbeit das Ziel, die gegenwärtigen Schwierigkeiten der praktizierten Diagnostik in der Schule aus dem Zusammenhang der behaupteten Begründungen zu erklären. Dabei geht es aber nicht um eine systematische Darstellung mehr oder weniger bekannter Einwände gegen die derzeitige Form der Schülerbeurteilung, sondern um eine methodologische und inhaltliche Begründung der Aussichtslosigkeit, die Selektionsdiagnostik methodisch optimieren zu wollen.

Hinter den Optimierungsversuchen im Bereich der Schülerbeurteilung steht der Anspruch, die "Objektivität" des Beurteilungsprozesses sicherzustellen. Mit Objektivierung ist dabei die Entsubjektivierung des Beurteilungsprozesses gemeint. Ob für die Erfüllung des empiristischen Prinzips der Entsubjektivierung im Bereich der Diagnostik in der Schule die notwendigen und hinreichenden Voraussetzungen vor allem im Hinblick auf die Existenz und zu erwartende Verlängerung einer auf die methodische Optimierung ausgerichteten Entwicklungslinie diagnostischer Forschung gegeben sind, wird im Mittelpunkt der Analyse stehen. Daß diese Entwicklungslinie auch innerhalb der BRD zunehmend verfolgt wird, zeigt bereits eine überblickhafte Betrachtung der Forschungstendenzen schulischer Diagnostik.

Die Beurteilung der Schüler zum Zweck ihrer Verteilung auf weiterführende Bildungs- und Ausbildungszweige vollzog sich bis zum Ende der 60er Jahre weitgehend im Rahmen der konventionellen Zensurengebung; sie wurde vom unbegründeten Vertrauen in die Zuverlässigkeit des Lehrerurteils und von der unterstellten Einsicht der entscheidungsbefugten Erziehungsberechtigten in die Gültigkeit solcher Urteile getragen. Zu einer Problematisierung bestand solange kein Anlaß, wie die Qualifizierungseffektivität des Bildungssystems unangetastet bleiben konnte und im Einklang mit der jeweils sich durchsetzenden Nachfrage nach den vermittelten Qualifikationen stand. Erst mit der PICTschen Ausrufung der

"deutschen Bildungskatastrophe" und der damit verbundenen Forderung nach Ausschöpfung der Begabungsreserven kamen ernstgenommene Zweifel über die Aussagefähigkeit des Lehrerurteils auf. In der Folge erschien eine Fülle von wissenschaftlichen Auseinandersetzungen mit der Zensurengebung (KÖTTER, 1965; TENT, 1969; GAUDE & TESCHNER, 1970; INGENKAMP, 1971; ULICH & MERTENS, 1973; PROJEKTGRUPPE, 1973; u.a.).

In diesem Zusammenhang ist die Tatsache wesentlich, daß sich die extensive Kritik am Lehrerurteil erst mit der Herausbildung eines Bewußtseinsprozesses entzündete, in dem die Beziehung zwischen Bildungssystem und gesellschaftlicher Entwicklung bis in ihre ökonomischen Auswirkungen erkannt wurde (COOMBS, 1967). Obwohl die Kritik an der Diagnostik dieser Bewußtseinsentfaltung nicht bedurft hätte - war doch aus der sozialpsychologischen Forschung die Problematik der Personwahrnehmung, sowie aus der psychodiagnostischen Forschung die Schwierigkeit der Eignungsbeurteilung zur Genüge bekannt -, ist die explizite Diskussion der gesellschaftlichen Relevanz von Bildung und Ausbildung für die Fortentwicklung der Diagnostik in der Schule von Bedeutung. Denn durch sie wurde nicht nur die durch die Subjektivität des Lehrerurteils bedingte Benachteiligung konkreter Schüler erneut thematisiert bzw. die Tatsache der Stabilisierung der Schichtenbenachteiligung durch die Zensuren in das Bewußtsein der Öffentlichkeit gerückt. In ihr wurde auch die grundsätzliche Frage nach der Aussagefähigkeit von Notenurteilen in Bezug auf die Kennzeichnung von Qualifikationsstrukturen gestellt und nach der Rückwirkung des Zensierungssystems auf das gesamte Bildungswesen, dessen Reformbedürftigkeit zumindest nicht zu bezweifeln und dessen konstitutiver Bestandteil das Notensystem war.

Aufgrund des als desolat eingeschätzten Zustandes des Bildungssystems wurden seinerzeit Reformmaßnahmen als nötig erachtet, die sowohl den durch die veränderte Arbeitsplatzstruktur bedingten Qualifikationsanforderungen an das Bildungssystem entgegenkommen sollten, als auch der veränderten Nachfrage nach Ausbildungsplätzen, verursacht durch die begonnene Mobilisierung der Bildungsmotivation breiter Schichten, entsprechen mußten. All-

gemein stand der Reformtrend in Zusammenhang mit den insbesondere durch Wirtschaft und Industrie vertretenen Interessen an einer verbesserten und dem technisch-industriellen Fortschritt angepaßten, wirklichkeitsadäquateren Durchschnittsqualifizierung der Auszubildenden (vgl. BAETHGE, 1970) und der von gesellschaftskritischer Seite vertretenen Forderung nach Abbau der Ungleichheit von Bildungschancen (vgl. PREUSS, 1973).

Soweit die zu verbessernde Durchschnittsqualifizierung betroffen war, wurde in den Reformmaßnahmen der Aspekt der gesellschaftlichen Relevanz von Bildung und Ausbildung aufzugreifen und zu realisieren versucht. Die Forderung nach Wirklichkeitsadäquanz der Inhalte war symptomatischer Ausdruck der empfundenen Notwendigkeit, institutionalisierte Bildung und gesellschaftlich durchgesetzte Ansprüche an sie zu reintegrieren. Auf die Neubestimmung, wenn auch nicht Erneuerung, der Unterrichtsinhalte richtet sich denn auch die in den vergangenen Jahren zunehmende Anstrengung im Bereich der Curriculumforschung.

Während sich die Forderung nach einer verbesserten Qualifizierung der Auszubildenden in organisatorischen Veränderungen und curricularen Reformen partiell konkretisiert hat (Gesamtschule, Orientierungsstufe, Förderstufe, Konzipierung von Curriculumstrategien, Lehrplanrevisionen u.ä.), vollzog sich eine davon getrennte Entwicklung im diagnostischen Bereich im Rahmen der konventionellen Selektionsfunktion des Bildungswesens und zwar vorwiegend als Optimierungsversuch der Schülerbeurteilung im Hinblick auf ihre schulerfolgsprognostizierende Effektivität. Diese Entwicklung wurde insofern begünstigt, als sich die realisierten Bildungsreformen, insbesondere die Gesamtschule, weitgehend als Reproduktionsformen der traditionellen Schulen erwiesen (HURRELMANN, 1971; EVERS, 1974).

Sie beginnt mit dem versuchsweisen Einsatz von allgemeinen psychologischen Tests (Intelligenztests, Motivationstests, allgemeine Leistungstests, Fragebogen zu neurotischen Reaktionstendenzen usw.), setzt sich fort in der Konstruktion von Schulreife und Schulleistungstests bzw. Auslese- und Übertrittstests

und mündet über die teilweise "unbefriedigend" ausfallenden Teacher-made-tests in dem Bemühen, sog. lernzielorientierte Schülerbeurteilungsverfahren herzustellen. Sein vorläufig letztes Stadium hat der Optimierungsprozeß in einem Versuch gefunden, der den bisher größten technischen Aufwand erforderlich machen könnte: der Konzeption einer Itembank. Bemühungen darum werden bereits heute organisatorisch repräsentiert durch eine Gruppe von Wissenschaftlern, die sich zu unregelmäßigen Tagungen unter dem Namen ITEM (Internationaler Erfahrungsaustausch über Test- und Evaluationsmethoden) zusammenfinden.

Der eher symptomatische Hinweis auf die Kontinuität und Hintergründe der Verwissenschaftlichungstendenzen in der Schülerbeurteilung verdeutlicht, daß die Versuche zur Objektivierung der praktizierten Diagnostik in der Schule nicht nur aus der Kritik am Lehrerurteil resultieren und nicht nur ein wissenschaftsimmanentes Problem darstellen. Die in der Kritik enthaltene Forderung nach präziseren diagnostischen Aussagen über die Qualifikation von Schülern und die entsprechenden Bewältigungsversuche von wissenschaftlicher Seite sind nur Erscheinungsformen für die Notwendigkeit, die Frage der Verteilung von Schülern auf weiterführende Bildungs- und Ausbildungszweige grundsätzlich zu überdenken. Gerade weil durch Urteile über Schüler konkrete Lebens- und Bildungskonsequenzen prognostisch festgelegt werden, ist die verwissenschaftlichte Urteilsbildung umsomehr auf ihre objektive Bedeutung für die Betroffenen zu hinterfragen.

1.2 Die Objektivitätsforderung als Konsequenz unbewältigter Probleme

Die kurze Skizzierung der Entwicklungslinie seit Beginn der 60er Jahre innerhalb des Bildungswesens der BRD könnte für sich genommen bereits ein erster Hinweis auf die Notwendigkeit einer Diagnostikreform grundsätzlicher Art in der Schule sein, zumal sich an der Zensierungspraktik im Prinzip nichts geändert hat. Erklärbar werden die Reformnotwendigkeit und die einzelnen Schritte in ihrer Folge nur dann, wenn man sie auf ihre funktionale Bestimmung bezieht: Mit dem Einsatz und der Konstruktion von standardisierten Testverfahren war die Zielsetzung verbunden, zuverlässige und gültige Prädiktoren (Eigenschaften und Verhaltenskomplexe von Schülern) zu finden, die eine optimale, d.h. möglichst fehlerfreie Verteilung der verschiedenen Schüler auf die weiterführenden Schulzweige gewährleisten. Dabei sollte ein Problem bewältigt werden, das an der traditionellen Zensur Umgebung noch am deutlichsten in Erscheinung tritt; es äußert sich in dem Versuch seiner testtheoretischen und testpraktischen Überwindung nur noch als methodisches, existiert gleichwohl als inhaltliches nachwievor. Es handelt sich um den Widerspruch zwischen der **m u l t i d i m e n s i o n a l e n** Determiniertheit des gesamten Schülerbeurteilungsprozesses (der Konstitution der Zensuren bzw. Testscores) und der **E i n d i m e n s i o n a l i t ä t** des Beurteilungsprozeßergebnisses (der Verwendung der Zensuren bzw. Testscores). Dieser Widerspruch wird letztlich durch die Funktionalisierung der Schülerbeurteilung für die institutionalisierten Selektionsprozesse innerhalb des bestehenden Bildungswesens erklärbar und damit solange als unlösbar ausgewiesen, wie sich das Ausbildungssystem durch den möglichst reibungslosen Vollzug der Selektion als effektiv präsentieren muß.

Ihren konsequenten Ausdruck findet die Funktionalisierung der Schülerbeurteilung in dem Verfahren der Diskrimination von Schülern, d.h., in dem Prinzip der maximalen Unterscheidbarkeit von Schülern auf der Ebene von Verhaltensdimensionen, deren poten-

tielle Lernerfolgsrelevanz als gesichert angenommen wird. Hinter diesem Prinzip verbirgt sich eine Beurteilungs-"logik", die durch die klassische Testtheorie systematisiert wird und sich in den Schwierigkeiten der inhaltlichen und empirischen Validitätsbestimmung von Prädiktoren des Schulerfolges, seien es Zensuren, Gutachten, Beurteilungsbögen oder Tests, manifestiert. Nach der klassischen Testtheorie muß ein plausibles Beurteilungsverfahren aus einer großen Anzahl von Aufgaben bestehen, die in Bezug auf die Schwierigkeit homogen sind und innerhalb der Anwendungspopulation maximal diskriminieren (BOCK & WOOD, 1971).

Welche Konsequenzen die Beurteilungslogik für die Interpretation der gemessenen - oder besser gesagt: der quasigemessenen - Dimension hat, wird im Verlauf der Arbeit noch eingehend untersucht werden müssen. Vorwegnehmend möchte ich darauf hinweisen, daß diese Beurteilungslogik ihre Existenzberechtigung ausschließlich durch die Selektionsfunktion des Bildungssystems erhält. Nur dadurch dürfte es zu erklären sein, daß man die Zensurengebung durch sog. objektive Tests zu ersetzen versucht, für die das Selektionsprinzip in seiner Erscheinungsform als Prinzip der maximalen Diskrimination konstitutiv ist, die deshalb an der Widersprüchlichkeit des Schülerbeurteilungsprozesses nichts ändern und die Schwierigkeit der Interpretation von Urteilen erneut offenlegen. Und nur dadurch dürfte zu erklären sein, daß trotz aller Kritik an der Zensurengebung die Note als Beurteilungsmerkmal nachwievor die zentrale Rolle spielt.

Unter dem Gesichtspunkt der Interpretationsschwierigkeiten läßt sich der in methodisch objektivierten Schülerbeurteilungen enthaltene Widerspruch folgendermaßen charakterisieren: Urteile haben im Hinblick auf die künftige Verwendung des Schülers eine konkrete, spezifische Bedeutung - Vergabe von Bildungs- bzw. Lebenschancen -, diese Bedeutung ist aber aus der schulischen Sozialisation des Betroffenen nicht ableitbar, weil in den Beurteilungen nur formale Distanzverhältnisse zwischen den Schülern zum Ausdruck kommen. Dies ist ein Merkmal aller Beurteilungen, die zum Zweck der Kalkulation des Erfolges in f e s t s t e -

h e n d e n Qualifikationsalternativen vorgenommen werden, für die also, weil sie organisatorisch und wurricular feststehen, eine Selektion durchgeführt werden muß.

Die Beurteilung von Schülern mit Ziffern, seien es Noten, Test-scores oder umgangssprachlich umschriebene Bezifferungen, stellt eine methodisch objektivierete Beurteilung dar, die sich durch folgende Merkmale auszeichnet:

1. Sie ist eine m e t h o d i s c h standardisierte Aussage über formale Distanzverhältnisse zwischen Schülern unter Zugrundelegung einer bestimmten Bezugspopulation; auch die Zensurengebung ist partiell methodisch standardisiert, z.B. durch die Anzahl der vorgeschriebenen Klassenarbeiten, die Darstellung des Notendurchschnitts als arithmetisches Mittel, die Empfehlung, Noten nach der Normalverteilung zu vergeben, o.ä.
2. Sie impliziert einen informellen Verbindlichkeitsanspruch in Bezug auf die Verwendbarkeit der Schüler in zur Verfügung stehenden Qualifikationsalternativen; gelegentlich ist dieser Verbindlichkeitsanspruch sogar gesetzlich abgesichert.
3. Sie bezieht den Verbindlichkeitsanspruch aus dem Anforderungsniveau der jeweiligen Qualifikationsalternative, das allein aus der postulierten Sachlogik des Curriculums ableitbar zu sein scheint, d.h., sie abstrahiert von der Subjektivität des Schülers und den situativen Umständen, die das Lernverhalten beeinflussen.

Daß die Bedeutung methodisch objektivierter Beurteilungen als Prädiktor von Bildungs- bzw. Lebenschancen für den k o n k r e t e n Schüler nicht auf seine Sozialisationsgeschichte - die schulische oder außerschulische - zu beziehen ist, belegt gerade die Fülle der empirischen Untersuchungen zur Bedeutungserhellung des Konstruktes "Schulleistung". Die Korrelationen zwischen einem ausgewählten Kriterium der Schulleistung (in der Regel doch immer wieder die Noten) und der Intelligenzdimension streuen je nach Bezugspopulation, nach Alter, nach Geschlecht, nach Meßinstrument, nach Selektionstechnik der Items usw. von Null bis ca. 0,80. Die Zahl der quasigemessenen Dimensionen, die mit der Schulleistung in Beziehung stehen oder könnten, ist grundsätzlich beliebig und de facto schon nicht mehr überschaubar, unabhängig davon, daß auch innerhalb der verschiedenen Dimensionen die Korrelationskoeffizienten stark streuen.

Hinzukommt, daß der Korrelationskoeffizient selbst ein stan--

dardisiertes Bezugsmaß ist, die Beziehung zwischen den quasigemessenen Dimensionen nur im statistischen Durchschnitt beschreiben kann und deshalb über den *I n d i v i d u a l f a l l* nichts aussagt. Einen eindrucksvollen Überblick zum Problem der empirischen Interpretation des Konstruktes "Schulleistung" liefert GAEDIKE (1974). Angesichts der Diffusität der Ergebnisse, für deren Integration nicht einmal ein Ansatz von Theorie zur Verfügung steht, kommt sie zu der Schlußfolgerung: "Die pädagogische Psychologie sollte von der bloßen Beschreibung (z.T. von Banalitäten) zur aktiven Einflußnahme auf die Vermittlung von Lernmaterial gelangen" (ebd., S. 87).

Die polemische Form dieser Kritik verdeckt ihren ernstzunehmenden Kern. Für die Optimierung der Schülerbeurteilung wird eine methodologisch systematisierende Forschungsarbeit herangezogen, die sich in Einführungen und Lehrbüchern zur mathematisch-statistischen Grundlegung schulischer Diagnostik und zur testtheoretischen Formalisierung pädagogisch-diagnostischer Fragestellungen darstellt. Man verwendet dabei einen scheinbar perfekten, methodischen Apparat zur Standardisierung der Schülerbeurteilung und beruft sich auf empiristisch-experimentelle Untersuchungen über den Zusammenhang von Schulerfolg und den ihn potentiell bedingenden Faktoren. Vergleicht man die methodologische Systematisierung mit der praktizierten Diagnostik in der Schule, deren Wesensmerkmal immer noch die Zensurengebung ist und die durch die Numerus-Clausus-Regelung paradoxerweise rehabilitiert wird, so wird man an die ironische Bemerkung eines Philosophen erinnert, die Realität doch endlich der Theorie anzupassen.

Es sei denn, man gelangt zu der Auffassung, daß die verschiedenen Theoriefragmente innerhalb der pädagogisch-psychologischen Diagnostik nichts anderes sind, als Beschreibungen von Teilaspekten des Prozesses der praktizierten Diagnostik in der Schule mittels formaler Modelle, deren beschränkte Adäquanz sich immer erst *n a c h t r ä g l i c h* und lediglich approximativ erweisen kann. Sicherlich könnten solche Modelle potentiell zu einer partiellen Erklärung der Wirkung bestimmter Einflußgrößen auf die praktizierten diagnostischen Tätigkeiten des beurtei-

lenden Lehrers beitragen; über die prognostische Relevanz des beurteilten Merkmals oder der beurteilten Ausprägung der modellimmanent beschriebenen Disposition des Schülers im Sinne der Zwecksetzung der Selektionsentscheidung wäre aber nichts auszusagen. Denn in der Anwendung des Modells wären seine Adäquanz erst zu belegen und damit seine Gültigkeitsbedingungen erst zu untersuchen.

Reduziert sich dabei das Adäquanzkriterium auf die Forderung, daß das diagnostische Modell die Unterscheidungsleistung der Zensurengebung zumindest zu reproduzieren habe - und dies in zuverlässiger Weise -, ohne daß es die Unzuverlässigkeit der Zensurengebung erklären oder die inhaltliche Dimensioniertheit der reproduzierten Unterscheidungen zwischen den Schülern ansatzweise aufklären könnte, dann verbleibt man mit diesem Modell unterhalb der Ebene der Feststellung von Banalitäten. Denn, daß sich Schüler im Resultat ihrer Auseinandersetzung mit der Situation "Unterricht", "Schule", "Ausbildung" usw. voneinander unterscheiden, ist jedem Lehrer und auch Schüler anschaulich in seiner Alltagserfahrung gegeben. Die konventionalisierte Prozedur der ziffernmäßigen Repräsentation dieser Unterschiede durch den Wertevorrat der Notenskala hat sich zum Zweck der auf die Selektion ausgerichteten Schülerbeurteilung im Durchschnitt der Fälle und im Hinblick auf die bestehende gesellschaftliche Funktion von institutionalisierter Ausbildung bis vor kurzem als "pragmatisch" hinreichend brauchbar erwiesen. Sie testtheoretisch rekonstruieren oder formalisieren zu wollen, muß unter dem pragmatischen Gesichtspunkt solange als überflüssig erscheinen, wie die darin enthaltene Irrationalität durch die Legitimation der Zensurengebung aufgrund ihrer historischen Tradition überdeckt werden kann.

Erst wenn die historische Tradition ihrer Legitimationsfunktion verlustig geht - und das ist eben dann der Fall, wenn die mit der Verteilung von Zensuren verbundene Vergabe von profilierten Qualifizierungschancen unter einen gesellschaftlichen Planungskalkül zu subsumieren versucht wird und unter den Zwang von

ökonomischen Rentabilitätsüberlegungen gerät -, stellt sich die Frage nach der "Objektivität" von diagnostischen Informationen über Schüler. Sie stellt sich, weil die Verteilung der Schüler auf kostenintensive Qualifizierungsalternativen bei gleichzeitiger Erwartung einer minimalen Irrtumsquote ansteht. Unter diesen Voraussetzungen ist allerdings zu untersuchen, welche Anforderungen an ein objektives Urteil des Lehrers über Schüler gestellt werden bzw. gestellt werden müssen, um entscheiden zu können, ob die Substitution der Zensurengebung durch einen methodisch objektivierten Beurteilungsprozeß die kritisierte Irrationalität des Lehrerurteils aufheben kann.

1.2.1 "Objektivität" als formalisierte Intersubjektivität

Beschränkt sich die Objektivitätsforderung auf die Notwendigkeit zur Entsubjektivierung und damit Standardisierung des Beurteilungsprozesses - ohne die Thematisierung des Zusammenhanges von inhaltlicher Dimensioniertheit des Urteils und der Möglichkeit von Entsubjektivierung zu beanspruchen -, so steht dabei im Hintergrund der Forderung ein formaler Begriff der Exaktheit wissenschaftlicher Aussagen. Objektiv ist demnach eine wissenschaftliche Aussage - hier ein Urteil über die Verhaltensstruktur von Schülern - genau dann, wenn sie unabhängig von Personen gilt, die diese Aussage treffen. Dieser Auffassung entspricht die Definition der Objektivität in der klassischen Testtheorie: "Ein Test ist objektiv, wenn er dasjenige Persönlichkeitsmerkmal, das er mißt, eindeutig mißt, d.h. wenn die Ergebnisse von der Person des Auswerterers unabhängig sind ..." (LIENERT, 1961, S. 13).

Durch den expliziten Verzicht auf die Berücksichtigung der metrischen Begriffsbildung und der Meßprozedur als objektivitätskonstitutive Momente des Schülerbeurteilungsprozesses bestimmt sich hier "Objektivität" ausschließlich als kontradiktorischer Gegensatz zu den intersubjektiv nicht eindeutigen Ergebnissen

der Schülerbeurteilung in Form der Zensurengebung. Sie impliziert somit keine Bedingungen für die inhaltlich-empirische Interpretierbarkeit der potentiell entsubjektivierbaren Beurteilungsprozesse.

Die Konsequenz daraus ist, daß die so im Verfahren der Standardisierung des Beurteilungsprozesses scheinbar eindeutig gemachten Urteile grundsätzlich beliebig interpretierbar sind und gerade deshalb de facto im Sinne der Selektionsentscheidungen interpretiert werden. Eine weitere Konsequenz ist die, daß die methodisch entsubjektivierten Urteile keine empirische Bedeutung haben müssen, da auch bloß konventionalisierte Verhaltenskriterien soweit "operationalisiert" werden können, um zwischen verschiedenen Beurteilern intersubjektive Übereinstimmung zu erhalten. Schließlich erlaubt die beschränkte Objektivitätsforderung die verfahrenstechnisch isolierte Überprüfung des Entsubjektivierungserfolges methodisch objektivierter Beurteilungsprozesse durch die testtheoretisch festgelegte Objektivitätsbestimmung, unabhängig von dem, was das Verfahren erfassen soll. Sie kann sich deshalb - und sie hat dies auch - zu einer *conditio-sine-qua-non* der Brauchbarkeit von Beurteilungsverfahren entwickeln, auch wenn mit ihr zirkuläre Bestimmungen der inhaltlichen Bedeutung dieser Verfahren in Kauf genommen werden müssen: Schulleistung ist, was der Schulleistungstest mißt.

In der Prioritierung der testtheoretisch systematisierten Objektivitätsforderung konkretisiert sich die Entwicklungstendenz innerhalb der empiristisch orientierten Sozialwissenschaften, sich an das Exaktheitsideal naturwissenschaftlicher Forschung anzulehnen bzw. an die durch die moderne Wissenschaftstheorie gelieferte Reflexion naturwissenschaftlichen Vorgehens. Nach diesem Ideal ist die Gewinnung von objektiv gültigen Aussagen zwangsläufig mit der Entsubjektivierung der Forschung in der Weise verbunden, daß - wie dies in der Formulierung Plancks am stärksten zum Ausdruck kommt - das Gesetz als vollkommen unabhängig von der Frage der menschlichen Fähigkeiten gelten muß. In der Tat finden in der naturwissenschaftlichen, konkreten Forschungsarbeit permanent derartige Entsubjektivierungsprozesse

statt; durch sie werden aber keinesfalls subjektive Erkenntnisprozesse schlechthin substituiert, sondern innerhalb angegebener Gültigkeitsbedingungen verallgemeinert und unter partiell konstruktivistische Modellannahmen subsumiert. Da sich in der naturwissenschaftlichen Forschung die so gemeinte Entsubjektivierung im Vollzug der wissenschaftlichen Analyse von Gegenstandsbereichen durchsetzt, ist eine unabhängig vom Wissenschaftsprozess durchzuführende Objektivitätsüberprüfung von Aussagen nicht mehr notwendig. Es existiert daher auch konsequenterweise kein methodisch formalisiertes Verfahren zur Überprüfung des intersubjektiven Gebrauchs von Aussagen, denn der intersubjektive Gebrauch ist eine Konsequenz der sprachlichen Rekonstruktion verallgemeinerter Analysen des Gegenstandsbereiches.

Im Gegensatz dazu ist in der Diagnostik in der Schule der intersubjektiv konsistente Gebrauch von Urteilen über Schüler eine notwendige Voraussetzung für die Analyse des Gegenstandsbereiches. Das zeigt sich u.a. daran, daß Validitätsüberprüfungen ohne die Einlösung dieser Voraussetzung nicht sinnvoll zu interpretieren sind. Dadurch gerät man in das Dilemma, zwar mit Begriffen operieren zu können, die den Anforderungen an die testtheoretische Objektivität genügen - z.B. mit dem durch einen Test festgelegten Begriff der Schulleistung -, über deren Bedeutung man aber nur statistische Hypothesen aufstellen kann, deren empiristische Überprüfung das gleiche Problem noch einmal aufwirft - wie im Fall der Validitätsbestimmung -, die Absicherung des intersubjektiven Gebrauchs des Kriteriumbegriffes. Die sich hier demonstrierende Umkehrung der naturwissenschaftlichen Methode der Entsubjektivierung wissenschaftlicher Theorienentwicklung und Theorienüberprüfung ist besonders hervorzuheben. Es kommt darin der ausschließlich paradigmatische Bezug der empiristisch orientierten Forschung im Bereich der Diagnostik in der Schule zur wissenschaftstheoretischen Reflexion naturwissenschaftlicher Forschungstätigkeit zum Ausdruck. Der dem übernommenen Paradigma zugrundeliegende, konkrete Forschungsprozeß bzw. seine wissenschaftstheoretisch reflektierte und verallgemeinerte Systematik wird nicht als Vergleichsebene heran-

gezogen.

Es scheint allerdings, daß in jüngster Zeit zumindest auf der Metaebene der Testtheorie die Diskussion der darin enthaltenen Problematik beginnt. Welche Konsequenzen sich daraus für die empiristisch orientierte Forschung als Ganzes ergeben würden, ist in diesem Stadium der Diskussion noch nicht abzusehen. Unzweideutig kann aber bereits jetzt der klassischen Testtheorie als Methodologie des diagnostischen Prozesses aus dem vorhandenen Problembewußtsein eine Absage erteilt werden. "Die klassische Testtheorie blieb ... schuldig, was ihr Name versprach. Als eine Theorie der psychologischen Messung kann die Anwendung der Korrelationsrechnung auf psychologische Testergebnisse nicht verstanden werden: Sie setzt nämlich, lediglich an der "Meßgenauigkeit" interessiert, die Definition und Existenz der zu messenden Dimensionen einfach voraus, sowie man bei der Beurteilung der Genauigkeit eines Thermometers natürlich (voraussetzen kann; der Verf.), daß die Temperatur ein bereits wohldefinierter physikalischer Begriff ist; ... Intelligenz etwa ist aber kein Begriff, welcher schon vor und unabhängig von ihrer Messung definiert wäre; vielmehr wird Intelligenz ja eben erst als der 'Leistungsgrad der psychischen Funktionen bei der Lösung neuer Probleme' ... verstanden, wobei der Leistungsgrad fraglos wiederum durch eine Messung, durch ein oder mehrere Testergebnisse erfaßt werden muß. Es fehlt gerade jene Theorie, welche 'Intelligenz' (oder spezifische Intelligenzdimensionen) definiert. Sieht man sich in der Psychologie nach derartigen Theorien um, so ist wohl die faktorenanalytische Intelligenzforschung jener Bereich, in welchem Intelligenzdimensionen relativ am klarsten definiert und beschrieben werden - doch beruhen diese Forschungen ihrerseits auf Intelligenztestergebnissen, womit der circulus vitiosus geschlossen ist" (FISCHER, 1974, S. 18).

In der Untersuchung der Bewußtlosigkeit der empiristisch orientierten Forscher im Bereich pädagogischer Diagnostik wie auch in anderen Bereichen eignungspsychologischer Diagnostik gegenüber dem erwähnten Teufelskreis scheint mir ein zentraler Ansatzpunkt

für die Aufklärung der Schwierigkeiten in der praktizierten Diagnostik in der Schule zu liegen. Diese Schwierigkeiten hängen mit der fälschlicherweise prioritierten Forderung zusammen, den intersubjektiven Gebrauch von Urteilen über Schüler vor und unabhängig von der inhaltlichen und damit auch empirischen Analyse des diagnostischen Prozesses abzusichern. Die Betonung liegt dabei auf dem Wort "Prioritierung", da sich darin die Umkehrung der wissenschaftstheoretisch reflektierten Methode naturwissenschaftlichen Vorgehens kennzeichnen läßt.

In den Naturwissenschaften stellt sich die Entsubjektivierung der Theorieentwicklung als Resultat einer zunehmend instrumentell verallgemeinerten Auseinandersetzung mit den verschiedenen Gegenstandsbereichen dar. Die Möglichkeit und Notwendigkeit der Entsubjektivierung basiert auf empirischen Operationen und Prozessen und die paradigmatische Funktionalisierung des Forschungsprozesses im Sinn der Planckschen Formulierung wird erst nach einer langen Entwicklungsreihe naturwissenschaftlicher Tätigkeit zu Beginn des 20sten Jahrhunderts programmatisch wirksam. Dagegen bestimmt sich die empiristisch orientierte Sozialwissenschaft im allgemeinen und die verwissenschaftlichte Diagnostik in der Schule im besonderen von allem Anfang an durch die Verpflichtung auf das Entsubjektivierungsparadigma und auf dessen formalmethodische und also gegenstandsunbezogene Systematisierung und Instrumentalisierung. Diese These wird im Anschluß an diesen Abschnitt näher begründet. Ich bewege mich dabei bewußt auf der Ebene immanenter Kritik empiristisch orientierter Sozialwissenschaft; zum einen deshalb, weil die Praktizierung von Diagnostik in der Schule sich dem empiristischen Anspruch - insbesondere in Bezug auf das Objektivitätsproblem - verpflichtet fühlt, und zum anderen, weil die Interpretationsmöglichkeit bzw. die empirische Deutbarkeit von Urteilen über Schüler in methodisch objektivierter Form als Problem nur dann geklärt werden kann, wenn der darin implizierte empiristische Anspruch ernst genommen wird, d.h., wenn die diagnostischen Aussagen den wissenschaftstheoretisch geforderten Konstitutionskriterien für wissenschaftliche Aussagen genügen. Ob die Einlösung des Anspruchs grundsätzlich mög-

lich ist und ob sich aus der potentiellen Einlösbarkeit Veränderungen für die Behandlung des eingangs erwähnten Widerspruchs in der Schülerbeurteilung ergeben, wird im zweiten Kapitel diskutiert.

Zusammenfassend betrachtet, liefert die erste Interpretation der Forderung nach intersubjektiv konsistentem Gebrauch von wissenschaftlichen Aussagen einen ausschließlich formalisierten Objektivitätsbegriff. Die Priorisierung der Sicherstellung von Objektivität im Prozeß der Theorienbildung und -überprüfung führt im Rahmen des diagnostischen Forschungsprozesses unabhängig und vor der inhaltlichen Untersuchung des diagnostischen Prozesses zur Beläufigkeit der Interpretation von Urteilen über Schüler. Eben deswegen kann nicht behauptet werden, daß methodisch objektivierete Urteile keine empirische Bedeutung hätten, sondern, daß man über ihren empirischen Bedeutungsgehalt keine Entscheidung treffen kann.

Letzteres stellt aber das Schlüsselproblem der Schülerbeurteilung dar. Solange es nicht gelingt, den empirischen Bedeutungsgehalt von Urteilen über Schüler auszuweisen, kann keine Form methodisch objektivierter Beurteilung die prognostische Effektivität der pragmatisch eingespielten und kritisierten Zensur Umgebung übertreffen. Denn, entsprechend der Zielsetzung der praktizierten Diagnostik in der Schule sollen Schüler in Bezug auf Ausprägungen von dispositionellen Eigenschaften beurteilt werden, deren Vorhandenheit als notwendige Voraussetzung für eine mehr oder weniger langfristige Prognose des Lernerfolgs in künftigen Qualifizierungsalternativen - seien dies die weiterführenden Bildungszweige im dreiklassigen Schulsystem, Differenzierungskurse innerhalb einzelner Schulen oder anschließende Jahrgangsklassen - angenommen wird.

1.2.2 "Objektivität" als Übereinstimmung im Gebrauch einer Beobachtungssprache

Wenn man von der diagnostisch hochgradig problematischen Frage vorerst absieht, wie sich im Bildungssystem festgelegte Lernziel- oder Qualifikationsanforderungen in Annahmen über lernerfolgsbezogene Eigenschaftsbestimmungen transformieren lassen, dann steht man als Diagnostiker vor der Aufgabe, nach welchen Kriterien die verschiedenen Resultate der Auseinandersetzung des Schülers mit dem Unterricht, also die verschiedenen Ergebnisse seiner Lerntätigkeit, taxiert werden sollen. Innerhalb der Zielsetzung, prognostisch effektive Schülerbeurteilungen durchführen zu wollen, erfordert die Klassifikation der Ergebnisse die Berücksichtigung von zwei Aspekten:

1. Die Resultate sind als Ausdruck bestimmbarer, empirisch dispositioneller Strukturen und Prozesse im Schüler zu interpretieren.
2. Die prognostische Relevanz der so bestimmten Dispositionen ist auf die zur Verfügung stehenden Qualifizierungsalternativen zu beziehen und aus diesen zu begründen.

Vor dem Hintergrund dieser beiden Aspekte ergibt sich für das Problem des intersubjektiv konsistenten Gebrauchs von Urteilsaussagen über Schüler eine veränderte Sichtweise. Es geht hier nicht mehr um die Sicherstellung des intersubjektiven Gebrauchs und um die Entwicklung einer Methode zu seiner Überprüfung schlechthin; dieses wäre ohne Schwierigkeiten durch eine optimale Standardisierung von Beurteilungsverfahren und durch eine hinreichend reglementierte "Interpretation" der Ergebnisse des standardisierten Beurteilungsprozesses zu erreichen.

Hier geht es zunächst um die Möglichkeit der Entsubjektivierung von Aussagen über Lernoperationen des Schülers, die dieser im Sinne des zweiten Aspektes innerhalb der jeweiligen aufnehmenden Qualifizierungsalternative auszuführen und zu beherrschen hätte. Ihre lerntheoriebezogene Verallgemeinerung könnte die Grundlage für die Begründung von prognostisch relevanten und zu beurtei-

lenden Dispositionen sein. Objektiv wären derartige Aussagen insofern, als sich die dadurch festgelegten Lernoperationen gesellschaftspolitisch intersubjektiv legitimieren und in ihrer sequentiellen Anordnung im Verlauf des organisierten Lernens in der jeweiligen Qualifizierungsalternative sich intersubjektiv begründen ließe. Die Form solcher Begründungen wäre etwa den Versuchen zur Lernzieltaxonomierung (BLOOM, 1956; GAGNE, 1969) ähnlich, wonach die Beherrschung einer bestimmten Lernoperation die notwendige Voraussetzung für die Ausführung und Beherrschung einer nachfolgenden sein soll ¹⁾.

Mit der Forderung nach der gesellschaftspolitischen Begründung der Auswahl zu vermittelnder Lernoperationen im organisierten Lernen und nach dem Aufweis der entwicklungs- und lernpsychologischen Relevanz ihrer sequentiellen Anordnung sind zwei zentrale Momente eines theoretischen Bezugsrahmens für die Diagnostik in der Schule benannt. Ihre Konkretisierung könnte Resultat einer ernst genommenen, inhaltlichen Curriculumreform und inhaltsbezogenen Curriculumevaluation sein. Verbleibt man auf der Ebene der empirismus-immanenten Betrachtung der Entsubjektivierung schulischer Beurteilungsprozesse, so auch und gerade hier die Existenz eines theoretischen Bezugsrahmens schulischen Lernens für die Entwicklung einer objektiven Diagnostik von Dispositionen vorausgesetzt werden. Der empirische Bedeutungsgehalt von Dispositionen, die den Charakter theoretischer Konstrukte haben, ist nicht durch willkürliche und pragmatische Operationalisierungen zu bestimmen. Eine bestimmte, pragmatisch nützliche Operationalisierung kommt nur in Bezug auf den intendierten Anwendungsbe-

1) Um naheliegenden Irrtümern vorzubeugen, möchte ich hervorheben, daß ich die "sequentielle Anordnung" nicht auf die zeitliche Organisation der Lehrinhalte beziehe, sondern auf die möglichen Formen der Auseinandersetzung der Schüler mit den Lehrinhalten. In der pädagogischen Literatur wird dafür häufig der Ausdruck "Qualifikation" bzw. "Lernziel" verwandt. Ich bevorzuge den Ausdruck "Lernoperation", weil dadurch sowohl der Bezug zum Lehrinhalt zum Ausdruck kommt, als auch der empirisch prozessuale Aspekt des Lernens einer bestimmten Auseinandersetzungsform mit dem Lehrinhalt bzw. dem Vermittlungsmodus "Unterricht".

reich der Theorie in Frage.

Sicherlich würde im Rahmen eines empiristischen Forschungsprogramms das objektivitätskonstitutive Moment der gesellschaftlichen Begründung der Auswahl zu vermittelnder Lernoperationen aus dem Bereich theoretischer Reflexion in den vorwissenschaftlichen des Entstehungszusammenhanges einer Theorie verwiesen werden (WESTMEYER, 1972); als notwendig zu lösendes Problem zum Zweck der Gewinnung von potentiell intersubjektiv gebrauchten, diagnostischen Aussagen über Lerndispositionen verbliebe aber noch die entwicklungs- und lerntheoretisch zu begründende Sequenzierung der zu vermittelnden und zu erwerbenden Lernoperationen im Vollzug des organisierten Lernens. Erst auf der Grundlage einer gewissen Vorerfahrung und Vorkenntnis der pädagogischen Adäquanz der sequenzierten Vermittlung und des Erwerbs von Lernoperationen und deren theoretischer Begründbarkeit wäre die Voraussetzung für die Hypostasierung noch zu bestimmender Dispositionen gegeben, die im Sinne der Selektions-"logik" als prognostisch relevante Schülermerkmale fungieren könnten.

Die beiden angesprochenen Momente eines theoretischen Bezugsrahmens für die selektionsorientierte Diagnostik in der Schule werden in der Reflexion und Ausübung des Diagnostizierens weder berücksichtigt noch konkretisiert. Deshalb ist die Kritisierbarkeit der Zensurengebung als intersubjektiv inkonsistent zugleich Ausdruck der theoretischen Unbegründetheit des diagnostischen Prozesses sowie seiner inhaltlichen Grundlage, des Lernprozesses im organisierten Lernen. Die intersubjektive Inkonsistenz der Zensurengebung ist daher nicht einfach auf den subjektivistischen Interpretationsspielraum des zensurierenden Lehrers zurückzuführen, sondern auf den objektiven Interpretationszwang, in dem sich der Lehrer befindet, wenn er komplexe Schülerverhaltensweisen nach angeblich prognostisch relevanten, aber theoretisch unbegründeten und pragmatisch vagen Kriterien des potentiellen Schulerfolges klassifizieren muß. Unter gleichen Bedingungen führt die zunehmende methodische Objektivierung des Schülerbeurteilungsprozesses durch Testverfahren oder restriktive Reglementierungen der Interpretation von Prüfungsergebnissen keineswegs zu einer

intersubjektiv konsistenten Klärung des Bedeutungsgehaltes von Urteilen über Schüler; sie bedingt aber eine Vereinheitlichung der Klassifikation von Schülerverhaltensweisen. In Ermangelung eines theoretischen Bezugsrahmens müssen dann die Klassifikationskriterien pragmatisch konventionalisiert werden.

Exkurs zur Ableitbarkeit von Lernoperationen

Der vorher genannte, zweite Aspekt, der bei der Klassifikation von Lernergebnissen zu berücksichtigen ist, beinhaltet die Forderung, die Taxierung der verschiedenen Resultate der Auseinandersetzung des Schülers mit dem Unterricht auf die zur Verfügung stehenden Qualifizierungsalternativen zu beziehen. Nur dann besteht die Aussicht auf die Rekonstruktion von prognostisch relevanten Dispositionen. Die oberflächliche Betrachtung dieser Forderung mag den Eindruck erwecken, die Entwicklung eines theoretischen Bezugsrahmens, den man als Vorleistung für einen empiristischen Ansatz in der Diagnostik betrachten könnte, sei grundsätzlich möglich und deshalb nur eine Frage des sukzessiven Erkenntnisfortschrittes in der Zeit. Die wissenschaftsmethodisch ausgerichteten Versuche einer rationalen Begründung der Sequenzierung von Qualifizierungsschritten bzw. Lernoperationen geraten aber in außerordentliche Schwierigkeiten, die die Frage nach der Adäquanz der Begründungsmodelle für die Curriculumentwicklung und Unterrichtsplanung aufwerfen. Dabei sind die unlösbaren Widersprüche, mit denen verschiedene Vorhaben und Ansätze der Planung von Qualifikationsproduktion und Qualifikationsvermittlung im bestehenden Gesellschaftssystem konfrontiert werden, noch nicht einmal berücksichtigt.

MEYER (1972) zeigt in seiner Analyse des Deduktionsproblems in der Curriculumforschung, daß deduktionistische Begründungsmodelle keine Voraussetzung für die Bestimmung von empirisch überprüfbaren Lernoperationen bzw. ihrer Sequenzierung sein können. Man kann von da aus der Argumentation von BLANKERTZ in der Einleitung zu diesem Buch zustimmen: "... erfahrungswissenschaftlich objektivierbare Deduktionshypothesen sagen nicht, welche curricularen Entscheidungen 'an sich' aus bestimmten Normen folgen, sondern sie protokollieren und prognostizieren, welche gesellschaftlichen Gruppen aufgrund welcher Normen welche Lernzielpräferenzen haben werden" (MEYER, 1972, S. 8).

Bei dieser Kritik ist aber zugleich zu berücksichtigen, daß sich hinter dem Begriff "erfahrungswissenschaftlich objektivierbare Deduktionshypothesen" in der Curriculumforschung eine Aussagenklasse verbirgt, die wissenschaftstheoretisch eine andere Struktur hat als jene, die durch den Namen festgelegt wird; gemeint sind damit Aussagen, die die Operationalisierung von Lernzielen repräsentieren sollen. Nun läßt sich aber einerseits zeigen, daß die im Vollzug der "Operationalisierung" hergestellten Beziehungen zwischen Lernzielen keine formallogischen Deduktionen sind, sondern Aussagen über erwartetes bzw. erwünschtes Schülerverhalten (MESSNER, 1970). Deshalb muß über die inhaltliche Eigenart

von Lernzielen durch Konvention bereits vorher entschieden sein und deshalb kann über die unter den Lernzielen zu subsumierenden Lernoperationen nur durch Konsensus entschieden werden. Zum anderen erweist sich die angeblich erfahrungswissenschaftliche Objektivierung der Hypothesen über das, was die Schüler in einer bestimmten Sequenz von Lernoperationen erwerben sollen, in praxi als eine mit Verhaltensbegriffen angereicherte Auflistung von Lehrstoffeinheiten. Sie ist keine Rückführung der hypothetischen Aussagen auf Aussagen der Beobachtungssprache über stattfindende Lernprozesse und auch keine partielle Zurückführung im Sinne der möglichen Bedeutungen von "empirischer Signifikanz" (vgl. STEGMÜLLER, 1970, Band II, Studienausgabe Teil B).

Der Einwand von BLANKERTZ gegen die rationale Begründbarkeit der Sequenzierung von Lernoperationen unter dem Gesichtspunkt der darin implizierten Normenproblematik gilt daher erst dann, wenn die entsprechenden hypothetischen Aussagen über den Erwerb von Lernoperationen nach wissenschaftstheoretischen Kriterien erfahrungswissenschaftlich objektiviert und deduktiv gewonnen werden könnten. Letzteres ist aber nicht nur beim gegenwärtigen Stand der Forschung und angesichts des praktizierten Vorgehens in der Curriculumentwicklung, sondern grundsätzlich zu bezweifeln.

Das Dilemma der rationalen Unbegründbarkeit ermöglicht und erzwingt die pragmatische Festlegung von Lernoperationen durch sachstrukturelle Gegebenheiten des Lehrstoffs. Es legitimiert darüberhinaus die naiv-empiristische Forschungsstrategie, normative Entscheidungen in den Entstehungsprozeß von Hypothesen abzuschieben und letztere nur noch "empirisch" zu untersuchen. Läßt man sich auf diese Forschungsstrategie ein, dann wird die Beschäftigung mit dem zu Beginn des Abschnittes 1.2.2 erwähnten ersten Aspektes ohne die Berücksichtigung des zweiten scheinbar möglich. Dort wurde gesagt, daß, über den zweiten Aspekt hinaus, sich die Resultate der Auseinandersetzung in Form von zu klassifizierenden und zu beurteilenden Schülerverhaltensweisen als Ausdruck bestimmbarer, empirisch dispositioneller Prozesse und Strukturen im Schüler interpretieren lassen müssen, damit diesen Dispositionen überhaupt prognostische Relevanz zugesprochen werden kann.

Um die dahinterstehende Problematik herausarbeiten zu können, ist zunächst eine allgemeine Kennzeichnung des Unterschiedes zwischen dem Begriff "empirisch" und dem öfters schon verwandten Begriff "empiristisch" durchzuführen. Es handelt sich dabei nicht um ei-

ne eindeutige Definition beider Begriffe, sondern um eine für die Vorklärung der Objektivitätsfrage notwendige Begriffsexplikation. Von der empirischen Bedeutung einer Aussage oder eines Begriffes soll dann gesprochen werden, wenn die Aussage oder der Begriff in einen theoretischen Bezugsrahmen integriert ist und wenn der Prozeß, wie beides auf beobachtungsmäßig Gegebenes bzw. erlebnismäßig Aufweisbares beziehbar ist, die Bedeutung der Aussage oder des Begriffes mitkonstituiert. Im Gegensatz dazu impliziert der Begriff "empiristisch" den Sachverhalt, daß eine Aussage oder ein Begriff der Beobachtungssprache bereits angehören oder, sofern dies nicht der Fall ist und beides empirische Bedeutung beanspruchen soll, die Aussage oder der Begriff mindestens partiell nach festgelegten formalen Regeln (Zuordnungsregeln) auf die Beobachtungssprache zurückführbar sein muß. Das wesentliche des Unterschiedes zwischen "empirisch" und "empiristisch" besteht darin, daß bei letzterem im Gegensatz zum ersten von der Existenz einer hinreichend brauchbaren Beobachtungssprache ausgegangen wird, deren Konstituiertheit außerhalb der Reflexionsbemühungen empiristisch-wissenschaftstheoretischer Analysen bleibt und für die empirische Interpretation von Aussagen und Begriffen als irrelevant angesehen wird. "Die Einwendungen gegen die verschiedenen Vorschläge zur Definition der empirischen Signifikanz sind vollkommen unabhängig davon, wie der dabei festgelegte Begriff der Beobachtbarkeit festgelegt wird" (STEGMULLER, 1970, Band II, Studienausgabe Teil B, S. 190).

Der Verzicht auf die Reflexion von "Beobachtbarkeit" und damit der Konstituiertheit dessen, was als beobachtungsmäßig Gegebenes erscheint, wird von STEGMULLER folgendermaßen und - wie ich meine - zirkulär begründet: "In allen experimentellen wissenschaftlichen Disziplinen müssen sich die Personen, die an diesem Forschungsprozeß teilnehmen, darüber einig sein, was beobachtbar ist und was nicht, also insbesondere darüber, ob zu gewissen Zeitpunkten und in bestimmten Gebieten - mit oder ohne Zuhilfenahme gewisser Instrumente - bestimmte Beobachtungen gemacht worden sind. Käme von einem bestimmten Zeitpunkt an zwi-

schen Experimentalphysikern, Molekularbiologen, Botanikern, Zoologen, Ethnologen Urgeschichtlern etc. **keine** Einigung mehr zustande, so würde die betreffende Disziplin entweder aufhören zu existieren oder doch zumindest solange stagnieren, bis eine entsprechende Einigung erzielt würde" (ebd., S. 189).

Daneben nimmt STEGMÜLLER die Behauptung, daß die Konstituiertheit des beobachtungsmäßig Gegebenen für die Beobachtungssprache unwesentlich sei, auch explizit zurück, da - wie er einräumt - die Bestimmung der "Beobachtbarkeit" in starkem Maße von Konventionen abhängt und deshalb mit Beobachtungssätzen kein Immunisierungspostulat verbunden zu sein braucht.

Unabhängig von der Haltbarkeit des empiristischen Verständnisses von Beobachtungssprache gilt es aber das Phänomen zu begreifen, warum eine zumindest scheinbar existierende Übereinkunft zwischen den Naturwissenschaftlern hinsichtlich des Beobachtbaren und erlebnismäßig Aufweisbaren und hinsichtlich dessen, was als Beobachtungssprache zu verwenden sei, besteht. Für die Analyse der Diagnostik in der Schule ist dieses Phänomen von besonderer Bedeutung. Es konkretisiert sich darin eine erste Stufe des naturwissenschaftlichen Paradigmas der als intersubjektiver Gebrauch von Aussagen aufgefaßten Objektivität.

Für die Abschätzung der Legitimität der Übernahme dieses Paradigmas als Geltungskriterium für diagnostische Aussagen ist es wichtig zu untersuchen, ob die Vernachlässigung der Konstituiertheit der "Beobachtungssprache" unter den gleichen Bedingungen wie in der Naturwissenschaft möglich ist. Auf die dafür wichtige Unterscheidung zwischen logisch-empiristischen und naiv-empiristischen Paradigma hat DICK (1974) hingewiesen. Es zeigt sich, daß das naturwissenschaftliche Vorgehen - hier bezogen auf die exakten Wissenschaften - dem logisch-empiristischen Paradigma zuzuordnen ist. Dort kann der intersubjektiv konsistente Gebrauch von Beobachtungsaussagen insofern Voraussetzung sein, als er die sprachliche Rekonstruktion einer empirisch instrumentellen Auseinandersetzung mit gesellschaftlich-historisch gegebenen Gegenstandsbereichen ist, also eine vorwissenschaftssprachliche, objektive Grundlage hat.

In der Diagnostik hingegen entbehrt die "Beobachtungssprache" einer solchen Grundlage, weil nicht die empirisch-instrumentelle Veränderung des Gegenstandes - hier der Schüler - selbst beschrieben oder beurteilt wird, sondern das Resultat seiner *i n t e n t i o n a l e n* Auseinandersetzung mit einer instrumentalisierten Situation. Da die Resultate ihren Konstitutionsprozeß nicht preisgeben und zudem nach Kriterien der Erwünschtheit klassifiziert werden, kann von einer empirisch bedeutungshaltigen Beobachtungssprache zur Beschreibung der Prüfungsergebnisse nicht mehr die Rede sein. Im Gegensatz zu den unter den formalen, testtheoretischen Objektivitätsbegriff subsumierten diagnostischen Aussagen, müßten sich Schülerurteile unter dem Gesichtspunkt der zweiten Form des Objektivitätsbegriffes auch auf die Konstitutionsprozesse von Prüfungsergebnissen im Schüler beziehen. Unter den Bedingungen der gegenwärtig praktizierten Diagnostik in der Schule läßt sich die Forderung nach dem intersubjektiv konsistenten Gebrauch von Urteilen über Schüler in Ermangelung einer empirisch bedeutungshaltigen Beobachtungssprache nicht einlösen.

In dieser Hinsicht unterscheiden sich auch die Modi der Schülerbeurteilung durch Zensurengebung und durch standardisierte Tests nicht voneinander. Daß man innerhalb der Forschung für dieses Problem zumindest im Ansatz sensibilisiert ist, belegt der Vorschlag von SUPER (vgl. WATSON, 1959), Objektivität soweit zu fassen, daß die Reizsituation "Test" für jeden Probanden die gleiche Bedeutung haben sollte. Es läge zunächst nahe, diesem Vorschlag dadurch entgegen zu kommen, daß man die Reizsituation Test soweit wie möglich standardisiert, um die verschiedenen subjektiven Interpretationen der Situation durch die Probanden einzuschränken. Die würde aber Kenntnisse darüber voraussetzen, welche Situationsaspekte von allen Versuchspersonen reaktionsadäquat interpretiert werden, unter welchen Bedingungen der Interpretationsprozeß als abgeschlossen betrachtet werden kann, damit die interindividuellen Unterschiede zwischen den Endreaktionen der Versuchspersonen auf die Testsituation im obigen Sinn als objektiv aufzufassen sind, und welche Situationsaspek-

te reaktionsinadäquate, subjektive Interpretationen provozieren.

Zur Untersuchung dieser Fragen wäre von Hypothesen auszugehen, in denen Zusammenhänge zwischen klassifizierten Situationsaspekten und klassifizierbaren Verhaltensweisen während des Prüfungsprozesses formuliert werden, und der Einfluß dieser Verhaltensweisen auf die Testergebnisse nachzuweisen. Beispielsweise könnte man die Hypothese untersuchen, daß die Anwesenheit eines unbekannten Versuchsleiters (Situationsaspekt) bestimmte Versuchspersonen verunsichert (das Verhalten, das beobachtungssprachlich zu konkretisieren ist) und dadurch zu geringeren Lösungshäufigkeiten führt, weil die Verunsicherung eine partielle Fehlinterpretation der Testaufgaben bedingt. Derartige Hypothesen, die für die pädagogisch-psychologische Forschung typisch sind, haben jedoch eine tautologische Struktur unter einer Bedingung: wenn eine von der hypothetischen Aussage unabhängige Merkmalsbestimmung für die Versuchspersonen unterbleibt, bei denen das angenommene Phänomen auftritt. Sie behaupten dann nichts anderes, als daß es Personen gibt, für die der unterstellte Zusammenhang zutrifft, bzw. andere, für die er nicht zutrifft. Wollte man aber den Bestimmungsversuch unternehmen, z.B. durch die Einführung einer intervenierenden Variablen "Angst", so würde man wiederum feststellen, daß es ängstliche Personen gibt und nichtängstliche, die die Hypothese bestätigen, usw. ad infinitum. Im konkreten Einzelfall wäre also nicht vorherzusagen, ob die Anwesenheit eines unbekannten Versuchsleiters sich auf den Lösungsprozeß ungünstig oder neutral auswirkt.

Die Schwierigkeiten, die sich für die diagnostische Forschung aus dem Aufgreifen der zweiten Form des Objektivitätsansatzes ergeben, sind hier nur angedeutet. Sie resultieren aus dem Versuch, die im Schüler ablaufenden Konsituationsprozesse von Prüfungsergebnissen empiristisch rekonstruieren zu wollen, ohne über die durch das logisch-empiristische Paradigma geforderten Beobachtungssätze für die Hypothesenbildung zu verfügen. Es könnte der Eindruck entstehen, daß die Berücksichtigung einer hinreichenden

Menge von Variablen im diagnostischen Forschungsprozeß eine wissenschaftlich begründbare Lösung des Rekonstruktionsproblems erbrächte. Diese Hoffnung erweist sich aber grundsätzlich als unberechtigt, da die erwähnten Forschungshypothesen sich durch die Besonderheit auszeichnen, einerseits einen Zusammenhang zwischen Aussagen über Situationklassen und Aussagen über Verhaltensklassen behaupten, andererseits die Behauptung für einen noch zu bestimmenden Individuenbereich gelten soll. DICK (1974) nennt solche Hypothesen unvollständige Hypothesen, weil sie bezüglich des offenen Individuenbereiches eine Leerstelle enthalten, und das dahinterstehende Forschungsparadigma das "naiv-empiristische". Unter dem naiv-empiristischen Paradigma wird zwar vom formalen Verfahren der empiristischen Hypothesenbestätigung Gebrauch gemacht, aber nicht, wie es dem logisch-empiristischen Paradigma entspräche, die Unterscheidung zwischen Entstehungszusammenhang und Geltungszusammenhang bei der Hypothesenbildung durchgeführt. Das hat zur Folge, daß die Entstehungsbedingungen der Hypothese unexpliziert und generalisiert in die Hypothesenbestätigung eingehen und die Annahme einer nicht abgrenzbaren Anzahl von Zusatzhypothesen z.B. in Form Aussagen über intervenierende Variablen erforderlich machen. Die Begrenzung der Anzahl ist dann eine pragmatische Entscheidung.

Zusammenfassend betrachtet reicht die zweite Bedeutung von Objektivität über die erste, ausschließlich formale hinaus. Die Forderung nach der Bestimmung der empirisch dispositionellen Strukturen und Prozesse im Schüler impliziert die Verfügung über eine Beobachtungssprache, die auf anschaulich Gegebenes verweist und nur deshalb intersubjektiv gebrauchten Urteilen über Schüler eine empirische Bedeutung verleihen kann. Diese Bedingung ist in der diagnostischen Forschung nicht erfüllt, da für die Rekonstruktion dispositioneller Prozesse und Strukturen nur die Prüfungsergebnisse vorliegen, nicht aber die Prozesse und Strukturen selbst.

1.2.3 "Objektivität" als intersubjektiv konsistente Deutung von invarianten Verhaltensmustern

Bei der systematischen Betrachtung der Erscheinungsformen des Objektivitätsproblems im Rahmen der praktizierten Diagnostik in der Schule habe ich bisher auf zwei objektivitätskonstitutive Momente der Schülerbeurteilung hingewiesen, die sich aus der Verpflichtung auf den empiristischen Forschungsansatz ergeben: den Theoretischen Bezugsrahmen für Aussagen über erfolgsrelevante Eigenschaften von Schülern und die Art der Beziehbarkeit solcher Aussagen auf beobachtungsmäßig Gegebenes bzw. auf Aussagen der Beobachtungssprache. Ersteres sollte die Beurteilbarkeit von Schülerverhaltensweisen als Resultate der Auseinandersetzung mit der Situation Unterricht unter Zugrundelegung gesellschaftlich durchgesetzter Qualifikationsbestimmungen normativ legitimieren und durch eine Theorie schulischen Lernens rational begründen. Durch letzteres wäre zu entscheiden, ob den Urteilen über Schüler darüberhinaus empirischer Bedeutungsgehalt zugeschrieben werden kann.

Berücksichtigt man den derzeitigen Forschungsstand im Bereich der Diagnostik in der Schule und die angedeuteten Zweifel an einer grundsätzlichen Einlösung des empiristischen Forschungsprogramms vor dem Hintergrund der Erfordernisse, die für die selektionsbezogene Diagnostik in der Schule bestehen, so läßt sich feststellen: Wir sind von der Möglichkeit eines intersubjektiv konsistenten Gebrauchs von Aussagen über Schüler in Form von Urteilen über deren langfristig erfolgsrelevante Qualifiziertheit für bestimmte Ausbildungsalternativen noch weit entfernt.

Dieses Dilemma ist von entscheidender Bedeutung für die Entwicklung einer Diagnostik in der Schule. Denn die beiden genannten Forderungen - die Vorausgesetztheit eines theoretischen Bezugsrahmens für Aussagen über erfolgsrelevante Eigenschaften von Schülern und des empirischen Bedeutungsgehalts solcher Aussagen - sollen die Grundlage für die Annahme von Dispositionen bilden,

d.h. von langfristig invarianten Schülermerkmalen, deren Erfolgsrelevanz eben durch diese Aussagen beurteilt wird. Ob beurteilte Dispositionen im Sinne der selektionsbezogenen Diagnostik funktional sind, wird deshalb davon abhängen, ob sich durch sie der "Schülerfolg" erklären läßt und ob sie durch Aussagen über beobachtbares Lernverhalten soweit partiell empirisch gedeutet werden können, daß dieses beobachtbare Lernverhalten als Kriterium für Selektionsentscheidungen heranzuziehen ist. In dem Maße wie die beiden Forderungen an die Bestimmung von Dispositionen nicht erfüllt sind, die verwandten Dispositionsbegriffe zur Unterscheidung von Schülern aber dennoch als auf empiristischem Weg in die Wissenschaftssprache eingeführt betrachtet werden, gerät man bei der Beurteilung von Schülern in Interpretationsschwierigkeiten. Ihre Beseitigung kann nur noch legitimatorisch erfolgen.

Das zentrale Problem stellt dabei die wissenschaftstheoretische Begründung der Annahme von langfristig invarianten Schülermerkmalen dar sowie ihre empiristische Rekonstruktion aus

1. den situativen Bedingungen, unter denen Schüler Prüfungsergebnisse erarbeiten - z.B. die Testsituation, die Situation, in der Klassenarbeiten geschrieben werden, die Situation, in der Schüler an sie gestellte Fragen beantworten sollen usw. -, und
2. den Prüfungsergebnissen bzw. den bereits beurteilten Prüfungsergebnissen.

Als Frage läßt sich das Problem folgendermaßen formulieren: Mit welchen Prüfungssituationen muß der Schüler konfrontiert werden und nach welchen Kriterien sind die Prüfungsergebnisse zu beurteilen, damit die beurteilten Prüfungsergebnisse als Indikatoren langfristig invarianter Schülermerkmale bzw. von Dispositionen betrachtet werden können ?

Für die Klärung dieser Frage ist eine kurze Skizzierung der wissenschaftstheoretischen Problematik von Dispositionsbegriffen und des darin implizierten Invarianzproblems erforderlich. Zuvor soll aber die Fragestellung an der selektionsorientierten

Schülerbeurteilung für die Selektionsalternative "Gymnasium" veranschaulicht werden.

Formal betrachtet, stellt die Selektion die Aufteilung einer Menge von Schülern in Teilmengen dar. Die in der Teilmenge zusammengefaßten Schüler zeichnen sich durch gemeinsame Bestimmungen aus: z.B. Schüler, die für die Hauptschule bzw. für das Gymnasium geeignet sind. Diese Bestimmungen beinhalten jedoch ein eigenartiges Charakteristikum; sie sollen sich nach einem gewissen Zeitraum nachträglich als realisierbar erweisen - am Beispiel des Gymnasiums nach 9 Jahren. Die Eigenartigkeit des Charakteristikums ist dadurch gegeben, daß das Kriterium der Einteilung der Schüler in Hauptschüler und Gymnasialschüler selbst noch als problematisch angesehen werden muß, weil über die Bedingungen der Sicherheit des Schulerfolgs in beiden Ausbildungszweigen soviel wie nichts bekannt ist, wenn man sich den E i n z e l f a l l vor Augen hält. Man verbindet zwar mit dem angewandten Einteilungskriterium eine gewisse Durchschnittserwartung in Bezug auf den möglichen Schulerfolg. Diese hat sich als eine Art Quasihochrechnung früherer Erfahrungen mit dem Einteilungskriterium durchgesetzt. Gleichzeitig geht aber in die Erfahrung die Gesamtheit aller "Fehlentscheidungen" ein, deren durchzuführende Analyse Bestandteil der Kritik am Einteilungskriterium sein müßte. Es sind die Fehlentscheidungen, die die Zulassung von "Ungeeigneten" betreffen, die am Gymnasium verbleiben bzw. es während der Ausbildungszeit verlassen, sowie Fehlentscheidungen, die die grundsätzlich unüberprüfbare Richtigkeit der Ablehnung von "Geeigneten" betreffen.

Angesichts der Unsicherheitsfaktoren in den globalisierten, früheren Erfahrungen und in den an das Einteilungskriterium geknüpften Durchschnittserwartungen ist die Selektionsentscheidung für den E i n z e l f a l l und, bemessen nach dem Anspruch einer optimalen Trefferquote, mit einer prophetischen Zukunftsvorhersage zu vergleichen; unabhängig davon, ob man als Einteilungskriterium Zensuren oder Testscores verwendet. Gleichwohl zwingt dieses Verfahren zu einer pädagogisch-psychologischen

A n n a h m e : Wenn man zum Zeitpunkt t eine Teilmenge von Schülern mit der Bestimmung "sie sollen sich zum Zeitpunkt $t+9$ Jahre als hochschulreif qualifiziert haben" bilden will, dan muß man die Annahme machen, daß in der Persönlichkeit des Schülers zum Zeitpunkt t strukturelle Bedingungen oder Gesetzmäßigkeiten vorherrschen, die über die besagte Zeit invariant bleiben oder durch die Ausbildung eine angebbare, spezifische Transformation bis zur Hochschulreife erfahren.

Wollte man die persönlichkeitsstrukturellen Bedingungen untersuchen und zur Grundlage der Schülerbeurteilung machen, so bräuchte man eine Theorie schulischen Lernens, mittels derer sich aus einer pädagogisch-psychologischen Qualifikationsbestimmung von Hochschulreife invariante Persönlichkeitsstrukturen ableiten ließen. Diese müßten zum Zeitpunkt t entweder auffindbar sein oder sie wären als transformierte Formen nachzuweisen. Da man aber weder über eine derartige Theorie schulischen Lernens verfügt und ihre Entwicklung wegen der Nichterfülltheit bestimmter empiristischer Prämissen nicht möglich ist, noch von einer notwendigen Qualifikationsbestimmung ausgehen kann (sie müßte im übrigen für die Selektionsalternative "Gymnasium" 9 Jahre gültig sein), ist dieser Ansatz zum Scheitern verurteilt.

Auf welche Weise läßt sich aber dann die existierende Annahme der intraindividuellen Invarianz von qualifikationsspezifischen Eigenschaften oder Dispositionen von Schülern scheinbar aufrechterhalten ?

Realiter wird gewissermaßen das Pferd vom Schwanz aufgezäumt und auch hier eine formale Umdrehung der empiristischen Methodologie vorgenommen. Ausgangspunkt ist die Tatsache, daß das Ergebnis der Auseinandersetzung des Schülers mit den Bedingungen der Qualifizierungsalternative (solche Bedingungen sind: Curriculum, Unterrichtsstil, Methodik, materielle Ausstattung, soziale Organisation der Klasse usw.) auf einer eindimensionalen Skala "abgebildet" wird - sei dies die Notenskala oder die Skala eines dafür entwickelten Tests. Obwohl das mit der Note bewertete Ergebnis der Auseinandersetzung, wie man aus der eigenen

Erfahrung weiß und durch die Forschung erwiesen wurde (FRIEDRICH & KÖHLER, 1975), multipel determiniert ist und komplexe Bedeutung hat, erscheint es durch den verwandten Bewertungsmaßstab nur noch als Ausprägungsgrad einer Disposition, im vorliegenden Beispiel der Disposition "Hochschulreife". Man denke etwa an die verschiedenen Bedeutungen der Durchschnittsnote 3 für die Qualifizierungsalternativen: Medizinstudium, wirtschaftswissenschaftliches Studium, Ingenieurausbildung, Offizierslaufbahn oder Beamtenlaufbahn im öffentlichen Dienst.

Diese Disposition ist aber inhaltlich nicht bestimmt, d.h., sie ist weder theoretisch abgeleitet, noch empiristisch durch Zuordnungsregeln im Sinne der Operationalisierung wenigstens partiell gedeutet. Sie wird im besten Fall nur intuitiv oder durch Analogiebildung festgelegt. Daher begründet sich der Schluß auf ihre qualifikationsrelevante Wirksamkeit nur durch die vorliegende interindividuelle Varianz der durch Ziffern gekennzeichneten Merkmalsausprägungen. M.a.W.: Der Beleg, daß die Disposition "Hochschulreife" vorliegt, scheint dann erbracht zu sein, wenn es im Vollzug des gesamten Schülerbeurteilungsprozesses gelingt, zwischen den Schülern zu diskriminieren, oder anders ausgedrückt, wenn es gelingt, eine varianzerzeugende Diagnostik durchzuführen. Infolgedessen scheint dann der Beleg für die intraindividuelle Invarianz dieser Disposition dann erbracht zu sein, wenn die Ergebnisse eines Beurteilungsprozesses zum Zeitpunkt t mit den Ergebnissen eines Beurteilungsprozesses zum Zeitpunkt $t+9$ Jahre korrelieren, d.h., wenn die, sich in den vorliegenden Varianzen ausdrückende, interindividuelle Struktur zu beiden Zeitpunkten annäherungsweise die gleiche ist oder sich die jeweiligen Rangordnungen entsprechen. In diesem Fall spricht man von der Validität eines Prädiktors.

Da bekanntermaßen ceteris paribus die Korrelation mit größer werdender Varianz zunimmt, muß unter der Voraussetzung der Selektionsfunktion der Beurteilung mit allen Mitteln dafür Sorge getragen werden, daß die Varianz nicht schrumpft, um die Annahme

der intraindividuellen Invarianz aufrechterhalten zu können. Die beiden Maßnahmen dazu sind die Standardisierung der Beurteilung (Entsubjektivierung des Beurteilungsprozesses) und die Standardisierung des Unterrichts (zunehmende Verplanung und Vereinheitlichung des Unterrichts und somit Entsubjektivierung des Vermittlungsprozesses). Die Standardisierung des Unterrichts ist die hinreichende Voraussetzung für die divergente und heterogene Entwicklung der Schüler ~~im~~ Verlauf des schulischen Lernens und die Standardisierung der Beurteilung garantiert die Streuung der Prüfungsergebnisse um einen Modalwert der Lernerfolgsausprägung.

Als wie immer konstituierter Ausdruck eines durchschnittlichen Lernerfolges von Schülern im Kontext eines bestimmten Unterrichts der Selektionsalternative "Gymnasium" könnte der Modalwert prinzipiell empiristische Voraussetzung für die Annahme sein, daß die Urteilsdimension eine Disposition im Schüler repräsentiert. Dazu müßte es allerdings gelingen, die schülerspezifischen Abweichungen vom Modalwert inhaltlich zu interpretieren und den Modalwert selbst theoretisch auf den Unterricht zu beziehen. Dieses Verfahren wäre aber unter den gegebenen Voraussetzungen tautologisch, denn

1. ist eine Lernerfolgstreue um einen Modalwert grundsätzlich immer feststellbar und damit inhaltlich wertlos; für die praktizierte Diagnostik in der Schule ist das Erzielen der Streue sogar notwendig konstitutiv;
2. könnte der Modalwert nur dann auf den Unterricht bezogen werden, wenn man die Prüfungsergebnisse aus dem Unterricht ~~e r - k l ä r e n~~ könnte, also eine Theorie schulischen Lernens hätte.

Ein auf diese Weise eingeführter Dispositionsbegriff kann zwar nachträglich - hier nach 9 Jahren Gymnasium - zu legitimieren versucht werden. Seine Erkenntnisfunktion bestünde darin, invariante und qualifikationsrelevante Verhaltensmuster entwicklungs-, lern- und persönlichkeits-theoretisch im Bezug auf das organisierte Lernen in der Schule zu begründen und die Bedingungen abzu-

grenzen, unter denen die invarianten Verhaltensmuster auftreten. In der Form der praktizierten Einführung liefert er aber keinerlei wissenschaftliche Begründung für die selektionsbezogene Prognose des Schulerfolges in organisierten Qualifizierungsalternativen. Es ist natürlich möglich, die Beurteilung von "Hochschulreife" durch bestimmte Verfahren soweit zu standardisieren und zu reglementieren, daß mehrere Beurteiler in der Vergabe von Ziffernurteilen intersubjektiv übereinstimmen und - was das wesentliche daran ist - die Korrelation dieser Urteile mit dem Einteilungskriterium zum Zeitpunkt t möglicherweise relativ hoch ausfällt. Dieser formalen Intersubjektivität in der Beurteilung steht aber die subjektivistische Beliebigkeit in der Interpretation des Lernerfolgsmodalwertes und der individuellen Abweichungen davon gegenüber.²⁾

-
- 2) Wie hilflos man in der Interpretation von Modalwerten bzw. von anderen Durchschnittswerten von Urteilsverteilungen ist, zeigt sich an der Bonus- und Malusvergabe im Rahmen des Numerus Clausus. Da die länderspezifisch verschiedenen Notendurchschnittswerte inhaltlich nicht auf die Ausbildung zu beziehen sind, die subjektivistische Eignungsinterpretation von Zensuren gegen rechtliche Grundsätze verstößt und die Nichtvergleichbarkeit von Zensuren zumindest auf der Ebene der Länderunterschiede zugegeben werden muß, wird "Objektivität" auf rechnerischem Wege hergestellt. Abgesehen von den dabei implizierten, ungelösten meßtheoretischen Problemen und den für die prognostische Relevanz von Zensuren wichtigen Veränderungen der Ausbildung in der Zeit, treten die Widersprüche dieses Verfahrens bereits dann in Erscheinung, wenn man die Unterschiede zwischen den verschiedenen Gymnasien innerhalb eines Landes in Auge faßt. Obwohl diese Unterschiede konkret offener zutage treten (Rekrutierungsbereich des Lehrpersonals und der Schüler, Kapazität des Lehrpersonals, materielle Ausstattung der Schulen, regional bedingte Anforderungsunterschiede zwischen den Schulen usw.) als die nur global faßbaren Unterschiede zwischen "typischen" Gymnasien verschiedener Länder, werden durch dieses Verfahren die Schulen eines Landes gleichgesetzt und die Schüler erhalten keinen schulspezifischen Notenbonus bzw. -malus. Überdies ist hinsichtlich der Interpretation der Notendurchschnittswerte kein rationaler Klärungsversuch in dem Streit erkennbar, ob die Unterschiede zwischen den Ländern eher auf verschiedene Zensierungsgewohnheiten zurückzuführen sind oder mehr auf qualitative Unterschiede in der Ausbildungseffektivität.

Es wird in der genaueren Analyse der praktizierten Diagnostik in der Schule zu zeigen sein, daß die bestehenden Schwierigkeiten bei der Gewinnung von objektiven Aussagen über lernerfolgsrelevante Dispositionen von Schülern nicht nur auf das Fehlen der objektivitätskonstitutiven Momente - des theoretischen Bezugsrahmens und der empirisch gehaltvollen Beobachtungssprache - zurückzuführen sind. Sie resultieren vor allem aus dem Vorgehen, die Annahme der intraindividuellen Invarianz von Verhaltenspotentialen aus der aufzuweisenden Stabilität interindividueller Varianzstrukturen methodisch zu rekonstruieren. Die Notwendigkeit, die Stabilität auf diese Art aufzuweisen, entspringt aber nicht mehr der Orientierung an einem naturwissenschaftlichen Paradigma bzw. an einem empiristisch wissenschaftstheoretischen Paradigma, sondern dem durch das bestehende Schulsystem gegebenen Selektionszwang. Er hat das für die praktizierte Diagnostik in der Schule konstitutive Diskriminanzprinzip zur Folge und läßt sich weder theoretisch begründen noch empiristisch bestätigen. Eine objektive, selektionsbezogene Diagnostik in der Schule dürfte daher nicht möglich sein.

In den Naturwissenschaften stellt sich das Dispositionsproblem und die Frage nach der Begründung der Invarianzannahme unter völlig anderem Gesichtspunkt. Ausgehend von dem beobachtungsmäßig gegebenen Phänomen, daß eine bestimmmbare Klasse von Gegenständen unter angebbaren Bedingungen ihren Zustand ändert und dieser Prozeß beobachtungssprachlich formuliert ist, soll eine sprachliche Form der Verallgemeinerung der Beobachtungsaussagen gewonnen werden, die die Änderungspotenz des Gegenstandes unter den bestimmten Bedingungen diesem als allgemeine Disposition zuordnet, auch wenn nicht jeder Gegenstand der Klasse der Prozedur unterworfen werden kann. Eine derartige Disposition ist z.B. "Wasserlöslichkeit" oder "Zerbrechlichkeit". Der wesentliche Unterschied gegenüber dem diagnostischen Dispositionsbegriff besteht darin, daß

- a) die invariante Reaktionstendenz der Gegenstände beobachtungsmäßig gegeben ist, wenn bestimmte Bedingungen beobachtungs-

mäßig gegeben sind;

- b) die Annahme der Invarianz der Zustandsänderung des Gegenstandes nicht über die graduelle Unterschiedlichkeit der Zustandsänderung verschiedener Gegenstände einer Klasse eingeführt wird, sondern über die qualitative Gleichheit des Prozesses bei allen betrachteten Gegenständen der Klasse.

Die Grundlage für die Einführung des Dispositionsbegriffes ist ein beobachtungsmäßig gegebener, empirischer Prozeß. Über seine beobachtungssprachliche Rekonstruktion kann man deshalb intersubjektiven Konsens herstellen, Von daher besteht das Problem nicht mehr in der inhaltlichen Klärung der Disposition. Es geht hier vielmehr um die formale Bestimmung der wissenschaftstheoretischen Struktur von Dispositionsbegriffen, die, weil sie nicht vollständig auf die Beobachtungssprache zurückführbar sind, in ihrer theoretischen Funktion ausgewiesen werden müssen. Die Frage nach dem intersubjektiv konsistenten Gebrauch von bereits bestehenden Aussagen über Unterschiede zwischen dispositionellen Eigenschaften von Gegenständen stellt sich deshalb im nachhinein nicht; solche Aussagen setzen schon den intersubjektiven Gebrauch der Beobachtungssprache voraus, durch die der Dispositionsbegriff partiell empirisch gedeutet wird.

In der diagnostischen Beurteilung von Dispositionen bei Schülern stellt sich allerdings erst durch die testtheoretische Objektivitätsüberprüfung heraus, ob überhaupt eine Chance besteht, deren empirische Bedeutung zu ermitteln. Darüberhinaus ist dies eine sehr fragwürdige Chance. Denn die Erscheinungsform des u.U. zufriedenstellenden, korrelativen Zusammenhanges zwischen dem intersubjektiv gebrauchten Einteilungskriterium für die Zulassung zur Selektionsalternative "Gymnasium" und der intersubjektiv gebrauchten Erfolgsbeurteilung am Ende des Gymnasiums sagt nicht mehr aus, als daß das Verhältnis zwischen den zugelassenen Schülern innerhalb des Zeitraumes von 9 Jahren relativ invariant geblieben ist. Dabei wird die Invarianz aus den Resultaten der Auseinandersetzung der Schüler mit zwei völlig ver-

schiedenen Prüfungssituationen extrapoliert.

Da man nicht angeben kann, in welcher Weise diese Verhältnis dimensioniert ist und inwieweit seine relative Stabilität auf den Einfluß durch das organisierte Lernen zurückgeführt werden kann, ist der durch das rein formal beschriebene, invariante Verhältnis legitimierte "Schluß" auf erfolgsrelevante Schülerdispositionen nur die nachträgliche Bestätigung dafür, daß bei der vorausgegangenen Selektionsentscheidung notwendige Zulassungsqualifikationen eine Rolle spielten. Diese inhaltlich zu explizieren scheitert daran, daß sie weder theoriebezogen noch empirisch eingeführt sind, d.h., daß sie im empiristischen Sinn nicht objektiv konstituiert sind. Gleichwohl bedarf das Phänomen eines feststellbar invarianten, inhaltlich jedoch nicht explizierbaren Verhältnisses zwischen den Schülern über den Zeitraum einer organisierten Lernphase hinweg einer Erklärung. Es wird an späterer Stelle zu zeigen sein, daß die relative Invarianz dieses Verhältnisses Ausdruck von langfristig invarianten und benennbaren Unterrichtsstandards ist.

In der zusammenfassenden Betrachtung der dritten Bedeutung von "Objektivität" als intersubjektiv konsistente Deutung von invarianten Verhaltensmustern zeigt sich, daß der Versuch, dispositionelle Eigenschaften von Schülern übereinstimmend zu beurteilen, nur noch pragmatisch legitimiert werden kann. Die Annahme der intraindividuellen Stabilität von lernerfolgsrelevanten Verhaltensmustern bzw. der Entwicklungsgesetzmäßigkeit von individuellen Lernverläufen läßt sich nur durch die Wiederholbarkeit eines Distanzverhältnisses zwischen den Schülern durch Prüfungsergebnisse aufrechterhalten. Es fehlt ihre wissenschaftliche Begründung durch theoretische Modelle und ihre empirische Bestätigung durch Befunde der pädagogisch-psychologischen Forschung. Der Hinweis auf das Fehlen dieser notwendigen objektivitätskonstitutiven Voraussetzungen im empiristischen Sinn wird aber erst dadurch erforderlich, daß hinter den Tendenzen zur Verwissenschaftlichung der Diagnostik in der Schule der empiristisch orientierte Objektivitätsanspruch steht.

1.3 "Objektivität" in der Diagnostik: Ausdruck der Entsub- jektivierung des Forschungsprozesses ohne Gegenstands- analyse

Bei der Diskussion der ersten Bedeutung von Objektivität habe ich die These vertreten, daß sich die empiristisch orientierte, pädagogisch-diagnostische Forschung durch die Verpflichtung auf das Entsubjektivierungsparadigma kennzeichnen läßt und seine formal-methodische und damit gegenstandsunbezogene Systematisierung und Instrumentalisierung den Forschungsprozeß leitet und bestimmt.

In diesem Abschnitt soll die Priorisierung dieses Vorgehens als Resultat der oberflächlichen Rezeption einer empiristischen Programmatik ausgewiesen werden. Das Ziel ist, den ausschließlich paradigmatischen Bezug der Diagnostik in der Schule zu den Erfordernissen einer empiristischen Forschung herauszuarbeiten. Dies soll aber erst im nächsten Abschnitt erfolgen. Um die Verwissenschaftlichungstendenzen im Bereich der schulischen Diagnostik zu verstehen, ist die legitimatorische Funktion des Entsubjektivierungsparadigmas für den Selektionszweck der Schülerbeurteilung aufzuzeigen. Dazu ist es notwendig, die Funktion von wissenschaftlichen Paradigmen zu erläutern und ihren Stellenwert im Forschungsprozeß zu umreißen.

1.3.1 Zwei exemplarische Funktionen des Entsubjektivierungs- paradigmas im Forschungsprozeß

"Es ist die Pflicht des Astronomen, die Geschichte der Himmelsbewegungen durch sorgfältige und geschickte Beobachtung zu erstellen. Dann muß er sich den Ursachen dieser Bewegungen zuwenden, oder vielmehr den Hypothesen über die Ursachen, die er sich ausdenken und entwerfen muß, weil er auf keine Weise zu den wahren Ursachen gelangen kann; und wenn man solche Hypothesen annimmt, kann man die Bewegungen nach den Prinzipien der Geometrie richtig berechnen, für die Zukunft ebenso wie für die Vergangenheit. Der gegenwärtige Autor hat diese beiden

Pflichten vorzüglich erfüllt. Denn diese Hypothesen brauchen weder wahr oder auch nur wahrscheinlich zu sein; wenn sie ein Rechenverfahren ergeben, das mit den Beobachtungen übereinstimmt, ist das allein schon hinreichend" (zit. n. TOULMIN, 1968, S. 49f).

Dieser Satz, der ebensogut von einem naiven Empiristen im 20. Jahrhundert hätte geschrieben werden können, stammt von einem protestantischen Theologen, namens Osiander, dessen Aufgabe es war, in einem Vorwort zu Kopernikus' "De revolutionibus orbium coelestium" der von kirchlicher Seite erwarteten Kritik den Wind aus den Segeln zu nehmen. Die Sätze sind für das vorliegende Problem von besonderer Bedeutung. Sie sind zum einen eine klassische Formulierung der prognostizistischen These, die in dieser Form in der modernen Wissenschaftstheorie nicht mehr aufrechterhalten wird, aber immer noch dem Prognosemodell der schulischen Diagnostik entspricht. Zum anderen wird in ihnen die Irrelevanz der Wahrheit oder Falschheit von Hypothesen gegenüber der Möglichkeit einer methodischen Rekonstruktion von beobachteten Ereignissen behauptet. Darin kommt eine relative Verselbständigung der formal-methodischen Gegenstandsbetrachtung gegenüber der inhaltlichen zum Ausdruck, die für die gegenwärtig praktizierte Diagnostik in der Schule prototypisch ist. Gleichzeitig sind aber die zitierten Aussagen deshalb bedeutsam, weil sie dem Anliegen von Kopernikus nicht nur nicht entsprechen, sondern dieses quasi auf den Kopf stellten. Im Gegensatz dazu interessierte sich Kopernikus - so jedenfalls TOULMIN - für eine Theorie hinter den methodisch darstellbaren Regelmäßigkeiten und wollte sich durchaus nicht mit einem Rechenverfahren zufrieden geben, das mit den Beobachtungen übereinstimmt.

Insoweit war er ein moderner Wissenschaftler, wenngleich sein Prinzip der "Regel der absoluten Bewegung" dem Mystischen verhaftet blieb. M.a.W.: Die Bedeutung dessen, was durch sorgfältige und geschickte Beobachtung vor Augen lag, war für ihn, im Gegensatz zur Rechtfertigungsargumentation von Osiander, nicht durch die rechnerische Synthese erschöpfend zu behandeln und forderte zur Untersuchung der zugrundeliegenden Ursachen auf.

Wenn Osiander die Realisierung eines methodischen Prinzips als die kopernikanische Leistung vorzustellen suchte, so ist dies vermutlich nicht nur darauf zurückzuführen, daß er Kopernikus gegen die Kritik durch die Theologen decken wollte. Immerhin wurde zu Beginn des 16. Jahrhunderts die Dringlichkeit einer Kalenderreform als allgemein notwendig angesehen, für deren Durchführung die kopernikanischen Forschungen auf dem Gebiet der Astronomie entscheidende Grundlagen lieferten. Da für diesen Verwendungszweck "ein Rechenverfahren ..., das mit den Beobachtungen übereinstimmt, ... allein schon hinreichend" ist, lassen sich mit dem Osianderschen Deckungsversuch zwei Fliegen auf einen Schlag treffen: die Kalenderreform kann auf gesicherten Erkenntnissen aufgebaut, die kirchlich-weltanschaulichen Grundlagen müssen nicht preisgegeben werden.

Ich besitze nicht die Kompetenz abzuschätzen, welchen Einfluß das Osiandersche Vorwort auf die Entwicklung der Naturwissenschaften bzw. auf die Entstehung eines naturwissenschaftlichen Methodenbewußtseins hatte. Es sollte an diesem Beispiel nur zum Ausdruck kommen, daß die Verselbständigung methodischer Prinzipien - die Forderung ein Rechenverfahren zu finden, das mit den Beobachtungen übereinstimmt - im Rahmen der Forschung auch zur Zeit von Kopernikus von Bedingungen abhing, die wir heute in außerwissenschaftlichen Bereichen zu lokalisieren gewohnt sind. Wenn ich vorher darauf hingewiesen habe, daß im Osianderschen Vorwort die prognostizistische These klassisch formuliert wird, so läßt sich damit zumindest verdeutlichen, daß die Verselbständigung methodischer Prinzipien sich im Hinblick auf Verwendungszwecke durchsetzt, die zwar auf die Absicherung einer Vorhersagetechnik angelegt sind - die Berechnung des Bewegungsverlaufes der Himmelskörper zum Zweck der Kalenderreform -, die aber für die Forschungstätigkeit im Sinne einer erkenntnisfunktionalen Prinzipien nicht konstitutiv sein müssen. Denn, "... eine neue und erfolgreiche wissenschaftliche Theorie führt unter Umständen zu keiner Verbesserung unserer Vorhersagetechniken, während andererseits erfolgreiche Vorhersagetechniken vielleicht

Jahrhunderte land ohne wissenschaftliches Fundament bleiben. Im ersten Fall braucht die Theorie darum nicht schlechter zu sein, im zweiten wird die Vorhersagetechnik nicht schon dadurch wissenschaftlich, daß sie funktioniert" (TOULMIN, 1968, S. 45).

Etwas über 400 Jahre später findet man eine dogmatisch anmutende Rechtfertigung der Erkenntnisfunktionalität verselbständigter methodischer Prinzipien im Bereich der Sozialwissenschaften; diesmal aber nicht in einem Vorwort, das vor vermutter Kritik bewahrensoll, sondern als Zielsetzung der gesamten Forschungsarbeit des Autors. "Diese Arbeit teilt in ihrem Grundansatz das Leitprinzip des Konstruktivismus ..., wonach wir nur das verstehen, was wir selber herstellen, also konstruieren können ... Es gehört zum Wesen eines konstruktivistischen Ansatzes, daß er sich nicht lange bei der Kritik des Gegebenen aufhält, sondern Normen einzuführen sucht, nach denen hier und heute gehandelt werden soll ..." (WESTMEYER, 1972, S. 20).

WESTMEYER versucht in seiner Grundlegung einer normativen Diagnostik so etwas wie eine "Logik der Diagnostik" zu entwickeln, "die dem Psychologen verbindlich vorschreibt, wie er sich in seiner Rolle als Diagnostiker zu verhalten hat, wenn er korrekt vorgehen will" (ebd., S. 20). Das Kriterium der Korrektheit ist die Grundlage der seinem Modell zugrundegelegten Logik, die als solche einer Begründung selbst nicht mehr bedarf und deren Adäquatheit als formale Methode der Analyse und Konstruktion von Aussagen über den diagnostischen Gegenstandsbereich vorausgesetzt werden kann.

Berücksichtigt man, daß WESTMEYER von der wissenschaftstheoretischen These der Strukturidentität von Erklärung und Prognose ausgeht und Diagnostik im weiteren Sinne als jenen Prozeß begreift, bei dem eine beobachtbare Verhaltensweise als vorliegendes Ereignis unter Bezugnahme auf allgemeine Gesetzhypothesen durch entsprechende Antecedensbedingungen erklärt werden soll, so läßt sich der Schlußsatz des zitierten Osianderschen Kommentars abgewandelt auch auf den diagnostischen Gegenstandsbereich übertragen: die zur Entwicklung herangezogenen Hypothesen brau-

chen weder wahr oder auch nur wahrscheinlich zu sein; wenn sie zu einem Verfahren der logischen Analyse und Konstruktion von Aussagen über beobachtbare Ereignisse führen, so ist das allein schon hinreichend. Berücksichtigt man außerdem, daß sich die sozialwissenschaftliche Forschung im allgemeinen und die im Bereich der Diagnostik im besonderen nicht nur logischer Methoden bedient, sondern auch auf mathematisch-statistischen Analysen als unentbehrlicher Grundlage aufbaut, dann kann das, was allein schon hinreichend zu sein scheint, entsprechend aus-
geweitet werden, ändert aber im Grundsätzlichen nichts.

Es bleibt allerdings die Frage offen, w o f ü r ein derartiges Verfahrensinventar hinreichend sein soll. Kopernikus jedenfalls sah in der Gewinnung von Rechenverfahren, die mit den Beobachtungen übereinstimmen, nicht das Ziel seiner Arbeit. Er war sich des Umstandes bewußt, daß "die Planetentheorien von Ptolémäus und der meisten anderen Astronomen mit den numerischen Gegebenheiten hinreichend übereinstimmen" (zit. n. TOULMIN, 1968, S. 51). Und Osiander mußte vermutlich aus den genannten Gründen den technologischen Aspekt von Kopernikus' Arbeit - die quantitative Bestimmung der Bewegung der Himmelskörper zum Zweck der Vorhersage - in den Vordergrund stellen, wenngleich dieser Aspekt nicht den Kern der Arbeit ausmachte und schon garnicht erkenntniskonstitutiv war.

Am Beispiel des Grundlegungsversuches einer normativen Diagnostik stellt sich die Ausgangslage mit verkehrten Vorzeichen dar: Das Ziel ist die Subsumption von Aussagen über beobachtbare Ereignisse unter eine logische Struktur, die selbst die Erklärung bzw. die Prognose darstellt und deren Konstruktionsregeln als Methodologie erkenntniskonstitutiv erscheinen. Nach WESTMEYERS konstruktivistischem Bekenntnis, wonach wir nur das verstehen, was wir selber herstellen, ist die Konstruktion des zu Verstehenden zwar nur notwendige und keine hinreichende Voraussetzung für das Verstehen. Von daher ginge er über die erkenntnistheoretische Problematisierungsebene in der Vorbemerkung von Osiander hinaus, der sich mit dem Zusammenhang von Hypothese und Rechenverfahren zufriedenzugeben scheint. Auf der anderen Seite erhebt

WESTMEYER mit seiner Entwicklung einer "Logik der Diagnostik" den Anspruch, ein Modell korrekter diagnostischer Vorgehensweise anbieten zu wollen; es ist somit die Vermutung nicht von der Hand zu weisen, für ihn sei mit der Konstruktion des zu Verstehenden, hinter der sich letztlich die Rekonstruktion eines Begründungszusammenhanges verbirgt, auch die hinreichende Bedingung für das Verstehen geschaffen, zumindest aber das Wesentliche für das Verstehen gegeben.

Letzteres weist darauf hin, daß für ihn die logisch-methodische Rekonstruktion des Begründungszusammenhanges erkenntnisfunktionale Bedeutung hat und mithin erkenntniskonstitutiv ist. Dagegen konnte die Methodik im Osianderschen Argumentationsvorgehen nur in Bezug auf bestimmte legitimatorische bzw. Verwendungszwecke als verselbständigte in Erscheinung treten und als Technik in den Dienst genommen werden. Sie war das formalisierbare Ergebnis einer Jahrhunderte langen Beschäftigung mit dem Gegenstandsbereich der Astronomie und entwickelte sich im Prozeß empirischer Erkenntnisgewinnung. Im Ansatz von WESTMEYER erhalten Methodik und Methodologie einen erkenntnisfunktionalen Stellenwert. Sie treten auch hier als verselbständigte in Erscheinung, allerdings vor dem empirischen Forschungsprozeß als dessen notwendige Voraussetzung im Sinne eines verbindlichen Regelsystems wissenschaftlichen Vorgehens.

Die Frage, für welchen legitimatorischen Zweck bzw. für welchen Bereich von Verwendungszwecken Methodik und Methodologie hinreichend wären, scheint hier nicht mehr stellbar zu sein, weil die Sätze der zugrundegelegten Logik als Teilmenge der Aussagen, die man in der Wissenschaftstheorie analytisch nennt, gegenüber der Erfahrung als immun betrachtet werden. Auf einen zusammenfassenden Nenner gebracht, repräsentiert die Immunitätsauffassung das Entsubjektivierungsprinzip in der Verwissenschaftlichung der Diagnostik in der Schule. Sie bietet dem einzelnen Forscher die Möglichkeit, sich auf allgemeine Regeln wissenschaftlichen Vorgehens zu berufen, deren Gültigkeit scheinbar unabhängig vom forschenden Subjekt vorausgesetzt werden kann.

Dies ist für den Bereich der diagnostischen Forschung umso problematischer, als die zur Anwendung kommenden methodologischen Prinzipien und Komplexe von methodischen Regeln nicht aus der Analyse der zu untersuchenden Gegenstände gewonnen wurden und sich für die Bewältigung diagnostischer Fragestellungen erst bewähren müssen. Deshalb ist mit der Entsubjektivierung des Wissenschaftsprozesses durch die Verpflichtung der Forscher auf die Verwendung angeblich gesicherter Methoden, die gegenüber der "Erfahrung" immun sein sollen, noch nichts erreicht.

Abgesehen davon, wird die Immunitätsauffassung in der neueren Wissenschaftstheorie in dieser eindeutigen Interpretation nicht mehr vertreten (STEGMÜLLER, 1970, Band II, Studienausgabe Teil B). Daß die Annahme der Immunität von analytischen Aussagen gegenüber der Erfahrung nicht aufrechtzuerhalten ist, manifestiert sich in der Entwicklung des modernen Empirismus in besonders hervortretender Weise, da man hier ursprünglich von der Grundüberzeugung ausging, die Gesamtheit der wissenschaftlichen Aussagen ließe sich in zwei dichotome Aussagenklassen aufteilen, die analytisch determinierten Sätze und die synthetischen Aussagen.

Wenn sich diese Annahme in der empiristisch-wissenschaftstheoretischen Diskussion als unhaltbar erwiesen hat, dann sollte sie von empiristisch orientierten Sozialwissenschaftlern nicht ohne entsprechendes Problembewußtsein rehabilitiert werden. Schließlich müssen sich gerade in den Sozialwissenschaften verschiedene methodische Ansätze erst legitimieren, und ihre Verwendung hat sich als pragmatisch brauchbar zu bewähren. Die Frage, für welchen legitimatorischen Zweck bzw. für welchen Bereich von Verwendungszwecken Methodik und Methodologie brauchbar scheinen, ist deshalb zu stellen. Für das methodische Prinzip der Entsubjektivierung innerhalb der Diagnostik in der Schule soll eine historische Betrachtung seines Wirksamwerdens zur Aufklärung des Verwendungskontextes im folgenden beitragen.

Gerade im Zusammenhang mit der Frage, nach der Möglichkeit eines intersubjektiv konsistenten Gebrauchs von diagnostischen Urteilen wäre es von grundsätzlicher Wichtigkeit, das Problem zu untersuchen, aufgrund welcher Bedingungen sich die Analytisch-Synthetisch-Dichotomie für lange Zeit halten konnte und sie als Paradigma wissenschaftlichen Vorgehens entstanden ist. Besondere Bedeutung haben dabei jene Bedingungen, die empiristisch orientierte Wissenschaftler gern in den Entstehungszusammenhang von wissenschaftlichen Aussagen abschieben. Durch eine historische Analyse des Entstehungszusammenhanges der Dichotomie ließe sich genauer klären, unter welchen spezifischen Zweckbestimmungen wissenschaftlicher Tätigkeit sich formallogische Maßstäbe für die Forschung als verselbständigte etablieren können.

Es ist nicht das Ziel dieser Arbeit, eine derartige Analyse durchzuführen. Ihre Realisierung würde den Rückgriff auf detailliertes, wissenschaftshistorisches Material erforderlich machen, dessen Aufbereitung innerhalb der Wissenschaftstheorie (vgl. KUHN, 1973; TOULMIN, 1968) und der Erkenntnistheorie (vgl. DAS ARGUMENT 88, 16. Jg., 1974, Heft 10-12) gerade erst begonnen hat. Die Exemplifizierung der beiden Formen der Verselbständigung wissenschaftlicher Methodik, zum einen im Dienst der Beibehaltung einer Ideologie und der Legitimation bestimmter Technologien, zum anderen als vorgefundenes und als verbindlich betrachtetes Regelsystem wissenschaftlichen Vorgehens, sollte die Untersuchung der Funktion von Methodik und Methodologie innerhalb eines Teilbereiches in den Sozialwissenschaften, der Diagnostik in der Schule, vorbereiten.

1.3.2 Historische Bedingungen für das Wirksamwerden des Entsubjektivierungsparadigmas im Bereich der Diagnostik in der Schule

Der für den Schülerbeurteilungsprozeß in Frage und zur Anwendung kommende Methodenkanon und seine dazugehörige methodologische Reflexion steht dem potentiellen Benutzer durch die vorherrschende wissenschaftliche Literatur als weitgehend verselbständigtes Verfahrensinventar zur Verfügung. Die Verselbständigung manifestiert sich z.B. in universitären, methodisch-propädeutischen Übungen, die vor aller inhaltlichen Auseinandersetzung mit dem diagnostischen Gegenstandsbereich zu Zulassungsbedingungen für die Fortsetzung des Studiums geworden sind. Sie manifestiert sich auch in kursusartigen Einführungen in die Technik der Testkonstruktion mit dem Anspruch, zu einer "objektiven" Beurteilung zu verhelfen. Sie konkretisiert sich schließlich in mathematisch-statistisch anspruchsvollen Abhandlungen über die Theorien des Testens ohne eine vergleichsweise annähernd intensive Behandlung der pädagogisch-psychologischen Probleme der Beurteilung.

Betrachtet man die immer noch vorhandene Euphorie, mit der an der Fortentwicklung der diagnostischen Methodik gearbeitet wird, so gewinnt man den Eindruck, die Betreffenden würden von der Abgeschlossenheit der inhaltlich-theoretischen Arbeit ausgehen können, ganz zu schweigen von der Gelöstheit der darin implizierten, gesellschaftlichen Probleme. Es sieht so aus, als stünde nur noch eine Aufgabe im Vordergrund: die ausschließlich methodische Bewältigung der Forderung nach "objektiver" Beurteilung.

Ich meine, daß diese vermittlungsorganisatorische Verselbständigung des methodologischen und methodischen Aspektes diagnostisch relevanter Forschung und Lehre spiegelbildlicher Ausdruck des Grundverhältnisses von theoretischer Reflexion und methodischen Optimierungsbemühungen im Bereich der Diagnostik ist.

Ganz allgemein scheinen die theoretischen Reflexionen des diagnostischen Prozesses innerhalb der Forschung und im besonderen innerhalb der praktizierten Diagnostik eine untergeordnete, wenn nicht nebensächliche Rolle zu spielen. Deutlich repräsentiert diese Tendenz das Handbuch der Psychologie (PSYCHOLOGISCHE DIAGNOSTIK, 1964), in dem, von wenigen Ausnahmen abgesehen, die Behandlung von methodischen Fragen und die Erörterung bereichsspezifischer, diagnostischer Verfahren und ihre Brauchbarkeit im Vordergrund stehen. Die Verlegenheit gegenüber theoretischen Problemen verrät sich auf eindrucksvolle Weise in den Auslassungen von BOESCH (1964) zur diagnostischen Systematisierung. Nach den Hinweisen auf die Notwendigkeit zu erkennen, daß man sich bei prognostischen Entscheidungen in den seltensten Fällen mit statistisch begründeten Funktionszusammenhängen zufrieden geben kann, und daß das Wesen diagnostischer Aussagen in prognostischen Möglichkeiten aufgrund funktioneller Einsichten besteht, beschließt er die Thematisierung dieser Frage mit der Bemerkung: "Die Ausführungen über die statistischen und funktionellen Normen müssen hier unvollständig bleiben. Sie seien nur Hinweise darauf, daß der Diagnostiker sich immer möglichst klar der statistischen wie funktionellen Bezugssysteme bewußt sein muß, auf die er seine Schlüsse aufbaut" (ebd., S. 935).

Ich meine weiterhin, daß das Mißverhältnis zwischen theoretischer Reflexion und methodischer Optimierung innerhalb des Forschungsgebietes "Diagnostik" und in weiten Bereichen der Sozialwissenschaften das Resultat der paradigmatischen Ausrichtung auf die Entsubjektivierungsforderung ist. In den Naturwissenschaften legitimiert die paradigmatische Formulierung dieser Forderung eine langfristige Forschungstradition und ein durchgängiges, wissenschaftliches Vorgehen und verallgemeinert dies auf einer wissenschaftstheoretischen Reflexionsebene. In der diagnostischen Forschung hat sie sich aber kurz vor Beginn des 20. Jahrhunderts ziemlich rasch in einem zunehmenden Methodenarsenal konkretisiert. Sie ist hier nicht die paradigmatische Verallgemeinerung tra-

dierter, konkreter Forschungstätigkeit, sondern repräsentiert die Erwartung, daß die formale Übertragung naturwissenschaftlicher, methodischer und verfahrenstechnischer Vorgehensweisen auf die Diagnostik zu entsprechenden Erfolgen bezüglich der Exaktheit wissenschaftlicher Analysen führen könnte.

Aufgrund des Fehlens einer empirisch-methodischen Forschungstradition war man bei den Versuchen, Diagnostik zu verwissenschaftlichen, auf Erfahrungen im Gebrauch von experimentellen Vorgehensweisen und ihrer methodischen Aufarbeitung angewiesen, wie sie in den Naturwissenschaften schon längst zum Forschungsalltag gehörten. Deshalb überrascht es keineswegs, daß die ersten "wissenschaftlichen" Psychologen, die sich mit diagnostischen Problemen auseinandersetzten, zugleich naturwissenschaftlicher Provenienz waren. "Diese Entwicklung beginnt insofern vor hundertvierzig Jahren, als die Wissenschaft von einer Erkundung des psychischen Individuums ein Abkömmling der modernen wissenschaftlichen Psychologie ist. Die Geburt dieser Wissenschaft aber ist bekanntlich eindeutig durch die Wendung zum Experiment gekennzeichnet. Die ersten Psychologen dieser Art sind ursprünglich P h y s i k e r oder auch M e d i z i n e r (Sperrung vom Verfasser). Sie lehnen sich in der Art ihrer Forschung aufs engste an das naturwissenschaftliche Experiment an. Die Grundbegriffe dieser Psychologie, wie etwa vor allem der Begriff der Empfindung werden im Laboratorium und anhand von Experimenten, welche dem naturwissenschaftlichen Denken entspringen, entwickelt" (HEISS, 1964, S. 6).

Die Übertragung empiristischer Methoden auf diagnostische Fragestellungen erfolgte keinesfalls immer unproblematisiert. Wenn gleich gegen Ende des vorigen Jahrhunderts im mitteleuropäischen Raum und in den USA noch wesentlich früher im Rahmen der Intelligenzforschung, der Persönlichkeitsforschung und auch in Bezug auf das Ausbildungswesen eine emsige Produktion von Beurteilungsverfahren einsetzte, waren die Übertragungsversuche sehr häufig von dem Bewußtsein getragen, das empiristische Vorgehen als Erprobungsfeld zu betrachten. Exemplarisch stehe hierfür RORSCHACH

(1921), der seine "Psychodiagnostik" noch als Methodik und Ergebnisse eines wahrnehmungsdiagnostischen Experimentes aufgefaßt wissen wollte.

Dennoch hat sich das Entsubjektivierungsparadigma gleichzeitig und in der Folgezeit soweit verselbständigt und zu einem Bewährungskriterium für Wissenschaftlichkeit entwickelt, daß sich sogar die traditionell antipsychotechnisch eingestellten und introspektionistisch orientierten Psychologiewissenschaftler tiefenpsychologischer Herkunft um die methodisch-experimentelle Untersuchung ihrer Begrifflichkeit und Aussagen bemühten (vgl. HÖRMANN, 1964). Die diese Entwicklungstendenz begünstigende Kritik an den subjektivistischen Interpretationsmöglichkeiten von projektiven Verfahren war auch ein Motiv für die Konstruktion und Verwendung von Beurteilungsverfahren im Ausbildungsbereich, obwohl sich der Ursachenkomplex für die Konstruktionsaktivität und das Verwendungsausmaß nicht allein dadurch bestimmen läßt. SÜLLWOLD (1964) konstatiert, daß sich ein genauer Zeitpunkt, von dem an Schultests verwendet werden, nicht angeben läßt. Seine Ausführungen deuten aber auf einen Beginn einer breiteren Entwicklungsaktivität um die Jahrhundertwende, wobei in Deutschland die Tendenz offensichtlich nach dem ersten Weltkrieg einsetzt.

Ungeachtet einer genaueren Datierungsmöglichkeit der Objektivierungstendenzen in der Schülerbeurteilung in Deutschland, fällt ihr relativ später Anfang im Vergleich zu anderen Ländern auf, zumal die auf das Subjektivismusargument abzielende Kritik am Lehrerurteil in Form der Zensurengebung schon wesentlich älter ist (vgl. DOHSE, 1967; ZIEGENSPECK, 1973). Man kann sogar behaupten, daß die Kritik an der Zensurengebung so alt wie diese selbst ist. So war die mangelnde Urteilsobjektivität einer der Gründe dafür, 1788 in Preußen mit dem Abiturientenexamen keine Berechtigungsfolgen zu verknüpfen, wenngleich nicht der wichtigste (zit. n. ZIEGENSPECK, 1973, S. 37).

Inwieweit die zeitliche Verschiebung im Beginn der Verwendung von objektivierten Schülerbeurteilungsverfahren in den einzel-

nen Ländern mit der Geschichte und Spezifität des jeweiligen Ausbildungssystems, mit dem staatlichen Einfluß auf die Ausbildung oder mit den gesellschaftlich durchgesetzten Ansprüchen an die Qualifizierung der Arbeitskraft zusammenhängt, kann an dieser Stelle nicht untersucht werden. Einer der wesentlichsten Gründe für die Bereitschaft, nach neuen Formen der Schülerbeurteilung zu suchen bzw. Ansätze zu neuen Formen überhaupt zur Kenntnis zu nehmen, dürfte in den allgemeinen Reformbewegungen im Bereich des Ausbildungswesens während der Weimarer Republik zu suchen sein. Damals schien "die Auflösung der Lehrrerseminare und die Hochschulmäßige Ausbildung auch der Volksschullehrer ... ebenso gesichert wie eine demokratische Reform des gesamten Schul- und Erziehungswesens überhaupt" (STÖHR, 1973, S. 171f).

Allerdings ist mit der damaligen, bildungspolitisch veränderten Situation der Hintergrund für potentielle Reformtendenzen nur global angedeutet. Wenn man diesbezüglich die praktizierte Diagnostik in der Schule betrachtet und das Wirksamwerden des Entsubjektivierungsparadigmas in Form objektivierter Beurteilungsverfahren verstehen will, so gilt es spezifische Bedingungen zu verdeutlichen, unter denen sich der Objektivierungstrend in der Schülerbeurteilung durchsetzen konnte. Ich glaube, daß hier drei wesentliche Bedingungen zu benennen sind, die dem Entsubjektivierungsparadigma zum Durchbruch verholfen haben und die die paradigmatische Funktion der Entsubjektivierungsmethoden als vorgefundenes und anzuwendendes Regelsystem wissenschaftlichen Vorgehens und als Legitimation für die Durchsetzung bestimmter Sozialtechnologien - hier der selektiven Verteilung der Schüler auf weiterführende Schulen - veranschaulichen.

1.3.2.1 Die psychotechnisch-diagnostische "Bewährungsprobe" des Entsubjektivierungsparadigmas in außerschulischen Bereichen

Um die Jahrhundertwende überschreitet die Entwicklung von objektivierten Beurteilungsverfahren den Rahmen der Intelligenzforschung und erstreckt sich auf die Bereiche der Industrie, Wirtschaft, Technik, Verwaltung und Bürokratie, sowie des Militärs. Legitimatorische Grundlagen für die spätere Übertragung auf den Ausbildungsbereich werden nicht nur durch die Testuntersuchungen für die amerikanische Armee im Jahre 1917 mit den "Group Examinations Alpha and Beta" geschaffen, sondern auch durch die Eignungsuntersuchungen für die Reichswehr nach dem ersten Weltkrieg und für die Wehrmacht unmittelbar vor dem zweiten Weltkrieg. Vergleicht man die verschiedenen Untersuchungen zur Entwicklung objektivierter Beurteilungsverfahren, so fällt der durchgehende Hinweis auf das Erprobungsfeld "Militär" auf, bei dem die massenhafte Anwendbarkeit solcher Verfahren ökonomischen Überlegungen entgegenkommt und ihre noch unüberprüfte inhaltliche Brauchbarkeit technologisch bestätigt. Unabhängig von ihrer möglicherweise aufzeigbaren Validität, rechtfertigen sich solche Verfahren allein schon durch ihren massenhaften Einsatz und durch die Erfüllung ihrer formalen Zweckbestimmung, der Unterscheidung von Individuen. Da die inhaltliche Klärung der Unterscheidungsdimension als grundsätzlich durchführbar angesehen wird, ist der Übertragungsversuch auf andere Anwendungsbereiche von anfang an legitimiert. So wurden z.B. Teile des Army-Test-Systems vielfach für die industrielle Personalauslese verwendet oder modifizierte Formen davon als Gruppentests für Schulkinder eingesetzt (vgl. GROFFMANN, 1964).

Solche Erfahrungen über die verfahrenstechnische Nützlichkeit methodisch objektivierter Beurteilungen in der Personalauslese für militärische Zwecke auf Massenbasis und der Qualifizierten-

selektion für bestimmte Beschäftigungsbereiche waren offensichtlich eine notwendige Voraussetzung dafür, daß der Widerstand der Lehrer gegen die Objektivierungstendenzen erschüttert werden konnte und das Entsubjektivierungsparadigma auch im unterrichtswissenschaftlichen Bereich eine pragmatisch relevante Funktion erhielt.

Auch in Europa lagen zunächst auf dem Gebiet der psychologisch beeinflussten Personalauslese und Berufsberatung einschlägige Erfahrungen über die potentielle Nützlichkeit methodisch objektivierter Beurteilungen vor (vgl. FELLENIUS, 1960), ehe dieses Vorgehen in der schulischen Diagnostik Eingang finden konnte und sich Arbeitsgemeinschaften von Lehrern und Psychologen nach dem ersten Weltkrieg mit dem Problem der objektivierten Auslese begabter Volksschüler befaßten (vgl. GROFFMANN, 1964). Der viel zitierte Einfluß der Forschungen von BINET und SIMON auf die Entsubjektivierungstendenzen in der Diagnostik ist in diesem Zusammenhang zu erwähnen, da sich deren Untersuchungen ursprünglich nur auf einen Teilaspekt der Schülerbeurteilung bezogen, nämlich auf die Beurteilung von schulisch rückständigen ("en retard") Kindern. Die damals schon längst bekannte Problematisierbarkeit der Schülerbeurteilung gab mithin nicht den Ausschlag für die Initiative des französischen Unterrichtsministeriums und für die wissenschaftliche Auseinandersetzung mit der interindividuellen Diskriminierbarkeit von Schülern.

Die aufkommenden Objektivierungstendenzen in der Diagnostik gegen Anfang dieses Jahrhunderts und ihre sukzessive Ausbreitung und Differenzierung in der nachfolgenden Zeit sind natürlich nur eine spezifische Erscheinungsform jener Rationalisierungsbestrebungen und Technologisierungsprozesse, die sich seit den 70er Jahren des vorigen Jahrhunderts im Zuge der vehementen Industrialisierung durchzusetzen beginnen und eine zunehmend geplante und gesteuerte Qualifikationserzeugung und -bewertung erforderlich machen. Erst mit der auf die Technologie abgestimmten, arbeitsteiligen Organisation der Produktionsabläufe war es möglich, Partialqualifikationen zu benennen, deren Vorhandenheit

oder graduelle Beherrschung die Frage nach ihrer objektivierten Beurteilung aufwerfen konnte. Die Umsetzung komplexer Arbeitsvollzüge in linear kombinierte, partielle und wiederholbare Arbeitstakte in der Produktion war eine objektive Voraussetzung für die methodisch objektivierter Betrachtungsweise menschlicher Tätigkeiten, nämlich als wiederholbare Tätigkeiten unter bestimmten, gleichbleibenden, materiellen Bedingungen. Deshalb erschien zumindest der Versuch plausibel, die Vorhersage einer gewissen Invarianz dieser Wiederholbarkeit durch das Verhalten in Prüfungssituationen zu diagnostizieren, die für die jeweiligen Tätigkeitsbereiche als repräsentativ angesehen wurden ³⁾.

Obwohl die in diesem Versuch enthaltene Problematik der Dispositionsbegriffe ungelöst bleibt, kann ein an wiederholbaren Tätigkeiten "validierbares" diagnostisches Verfahren insofern auf objektivierten Grundlagen aufbauen, als die Invarianz der Tätigkeiten zumindest durch die relative Unverändertheit der Arbeitssituation, der Arbeitsmittel und des herzustellenden Arbeitsproduktes festgelegt ist. Sicher ist eine durch repetitive Tätigkeiten festgelegte Arbeitsplatzsituation mit der Situation "Unterricht" in toto nicht vergleichbar. Bei letzterer kann die Bewältigung der diagnostischen Fragestellungen gerade nicht zu dem Versuch führen, durch irgendwelche, als repräsentativ betrachtete Prüfungsbedingungen langfristig invariante und auf die Wiederholbarkeit von repetitiven Tätigkeiten bezogene Dispositionen zu ermitteln. Allerdings sind die mit den Optimierungsansätzen zur methodisch objektivierten Beurteilung gesammelten Erfahrungen in Bezug auf die grundsätzliche Funktionalität eine notwendige Vorbedingung für die Übertragungs-

3) Eine detaillierte Untersuchung des Zusammenhanges von Qualifikationsbewegung und Entwicklung der Produktionsweise existiert m.E. bis heute nicht. Eine der Hauptursachen ist vermutlich darin zu sehen, daß die inhaltliche Struktur des Qualifikationsbegriffes selbst noch völlig offen ist und seine häufig implizit empiristische Interpretation als Reaktionspotential unter bestimmten Bedingungen im Sinne einer Disposition (Fähigkeit oder Fertigkeit) nicht erkannt wird (vgl. OTTEN, 1973).

absicht in den schulischen Bereich. Zwar war das Terrain für die Herreinnahme von methodisch objektivierten Verfahren in die Diagnostik in der Schule durch die Kritik an der Subjektivität des Lehrerurteils im Prinzip schon vorbereitet; um aber in das traditionell konservative Ausbildungswesen Eingang zu finden, bedurften sie der Bewährungsprobe in außerschulischen Bereichen.

1.3.2.2 Die Objektivierung der Kritik am Lehrerurteil durch das Entsubjektivierungsparadigma

Es wurde vorher schon darauf hingewiesen, daß die Kritik am Subjektivismus des Lehrerurteils durch Ziffernnoten letztlich mit der Einführung der Zensuren beginnt. Gleichwohl mußte diese Kritik solange wirkungslos bleiben, wie ihr selbst der Subjektivismusvorwurf gemacht werden konnte und die Legitimationsfunktion der Zensuren explizit nachweisbar war. Erst mit der Entwicklung von deskriptiv-statistischen Methoden zur Überprüfung der intersubjektiven Inkonsistenz des Lehrerurteils waren reglementierte Bedingungen dafür gegeben, die Kritik formal zu verallgemeinern. Durch die mathematisch-statistischen Methoden war sie soweit zu objektivieren, daß ihr gegenüber der Subjektivismusvorwurf nicht mehr gebrechtfertigt erschien.

Ich möchte an dieser Stelle die Betonung auf das Wort "erschien" legen, weil die formale Kritik keinerlei Ansatzmöglichkeiten liefert, den Schülerbeurteilungsprozeß inhaltlich auf eine objektive Grundlage zu stellen, wie das bisweilen aus Bemerkungen über die angebliche Objektivierungsleistung von Tests oder ähnlichen Beurteilungsinstrumenten hervorgeht. Der korrelationsstatistische Aufweis der intersubjektiven Inkonsistenz von Urteilen über Schüler belegt nicht mehr als den individuell unterschiedlichen Modus der Vergabe von Zensuren. Er erbringt als mathematisch-statistische Deskription der Vergleichbarkeit von Varianzstrukturen zweier Ziffernreihen keine Information über

die Ursache der relativen Stärke des Zusammenhangs beider Reihen. Der allgemeine Konsensus in Bezug auf die intersubjektive Inkonsistenz des Lehrerurteils ist damit nur in der Verwendung einer reglementierten - also entsubjektivierten - Überprüfungs-methode begründet, die zu ihrer Voraussetzung hat, daß Urteile über Schüler durch Ziffern repräsentiert werden. Da der Konstitutionsprozeß der Ziffernurteile für die Berechnung des korrelativen Zusammenhanges völlig irrelevant ist, läßt sich mit Hilfe des numerischen Korrelationskoeffizienten die Stärke des Zusammenhanges von Urteilen nicht erklären.

In den 20er und 30er Jahren hatten die Untersuchungen zur Uneinheitlichkeit der Zensierung (LIETZMANN, 1927; KIESSLING, 1929) einen wirksamen Einfluß auf die Objektivierungstendenzen im Bereich der Diagnostik in der Schule. Die offensichtliche Unbestechlichkeit der Überprüfungsverfahren zum Nachweis der subjektivistischen Vergabe von Zensuren stellt jene notwendige und kritische Vermittlungsebene dar, auf der sich die subjektivistische Willkür der Benotung methodisch "rein" konkretisierte. Durch sie fand das Entsubjektivierungsparadigma der naturwissenschaftlich orientierten Psychologie eine nachträgliche Bestätigung. Es bleibt festzuhalten, daß ersteres - die Objektivierung der Kritik am Lehrerurteil - wiederum nur, wie auch die Bewährung von instrumentellen Beurteilungsverfahren in außerschulischen Bereichen, eine weitere notwendige Voraussetzung für den Beginn der Entwicklung von objektivierten Beurteilungsverfahren in der Diagnostik in der Schule war. Durch sie konnte die Abhängigkeit des Urteils über Schüler vom Beurteiler "Lehrer" verbindlich nachgewiesen werden.

Daraus folgt keinesfalls zwangsläufig, den Beurteilungsprozeß als solchen insgesamt zu entsubjektivieren, wie dies der Struktur des testtheoretischen Objektivitätsbegriffes entspricht und in der praktizierten Testkonstruktion angestrebt wurde und wird. Aus dem formalen Nachweis der erwähnten Abhängigkeit folgt lediglich die Notwendigkeit, den Konstitutionsprozeß der Zensurenbildung zu analysieren. Dabei ist die Frage zu stellen, ob unter den Bedingungen des organisierten Lernens eine entsubjektivierte

Beurteilung in der Tat von subjektivistischen Komponenten frei sein kann, ob die Beurteilung von qualifikationsrelevanten Dispositionen möglich ist und ob entsubjektivierte Urteilsaussagen in dem Sinne objektiv sein können, daß ihnen empirische Bedeutung zukommt und sie in dieser Bedeutungshaftigkeit in einem theoretischen Bezugsrahmen stehen.

Mit der Objektivierung der Kritik am intersubjektiv inkonsistenten Lehrerurteil durch die korrelationsstatistische Beschreibung der fehlenden Übereinstimmung zwischen verschiedenen Beurteilern waren diese Fragestellungen noch nicht einmal angegangen. Immerhin konnte man aber mit derartigen Untersuchungen den Anspruch verbindlich widerlegen, die Zensurenhgebung würde zu einer gerechten und voraussetzungsadäquaten Beurteilung der Schüler führen. In der Funktion der objektivierten Darstellung, Urteile nicht intersubjektiv konsistent interpretieren zu können, bestand und besteht die essentielle Bedeutung von wissenschaftlichen Methoden zur Überprüfung der Eigenschaften von diagnostischen Verfahren wie Objektivität, Reliabilität, Validität usw. Eine derartige Überprüfung ist allerdings erst dann möglich, wenn Urteile in Formen vorliegen, die den Anschein von intersubjektiver Konsistenz erwecken. Sie ist ferner erst dann möglich, wenn gesellschaftlich durchgesetzte Ansprüche an den Nachweis von in der Schule erworbenen Qualifikationen Urteilsformen erforderlich machen, die das Berechtigungswesen eines selektionsorientierten Bildungssystems zu legitimieren in der Lage sind.

Letzteres ist für den Beginn der Entwicklung und Verwendung von objektivierten Verfahren im Bereich der Diagnostik in der Schule in Deutschland nach dem ersten Weltkrieg besonders wichtig, weil mit der veränderten wirtschaftlichen und industriellen Lage und der damit verbundenen Veränderung der Qualifikationsstruktur der Arbeitskräfte umfassende Schulreformen notwendig wurden. Sie bedeuteten die endgültige Überwindung des ständischen Ausbildungswesens (vgl. OTTEN, 1973). Die tradierte Zulassungsprozedur für weiterführende Schulen geriet unter beson-

deren Legitimationsdruck: Der zunehmende Bedarf an höherqualifizierten Schülern erzwang die qualitative Erweiterung der Rekrutierungsbasis für die weiterführenden Bildungszweige und die Durchsetzung von Auswahlkriterien für die Zulassung, nach denen zunehmend der erreichte Ausbildungsstand erfaßt werden mußte, soweit er für den langfristigen Verbleib in der höheren Bildungsalternative wesentlich zu sein schien.

Die ökonomisch und politisch bedingte Veränderung der Arbeitsplatzstruktur und die daraus resultierenden Anforderungen an die Qualifizierungsleistung des Ausbildungswesens sind letztlich die zentralen Ursachen für das sukzessive Wirksamwerden der Forderung nach Vergleichbarkeit der Schülerbeurteilung. Sie bilden den gesellschaftspolitischen Hintergrund für den Durchbruch des Entsubjektivierungsparadigmas innerhalb der Diagnostik in der Schule mit der Konsequenz, den Beurteilungsprozeß selbst zu entsubjektivieren. Zulassungsberechtigung für weiterführende Bildungszweige kann jetzt nicht mehr nur das ständisch verankerte Bildungsprivileg sein; an zunehmender Bedeutung gewinnt vielmehr das ausbildungsrelevante und feststellbare Qualifikationsniveau.

1.3.2.3 Vorläufige Zusammenfassung: Entsubjektivierung als formales Prinzip

Bevor ich auf die Frage der veränderten Qualifikationsanforderung als der dritten und entscheidenden Bedingung für die Entwicklung der Entsubjektivierung zum Forschungsparadigma näher eingehe, möchte ich die paradigmatische Funktion der Entsubjektivierungsforderung durch den Einfluß der zwei bisher genannten Bedingungen auf die Diagnostik in der Schule kennzeichnen - und zwar als gegenstandsunbezogenes Steuerungsprinzip der Verwissenschaftlichung der Schülerbeurteilung.

Im ersten Fall ging von der psychotechnisch-diagnostischen Bewährung der Entsubjektivierung von Beurteilungsprozessen ein

z.T. direkter Einfluß auf die Diagnostik in der Schule aus. Hier manifestiert sich die Gegenstandsunbezogenheit bei der Übertragung der methodischen Objektivierung der Beurteilung in der Ignoranz gegenüber der Tatsache, daß die diagnostischen Instrumente für einen spezifischen und begrenzten Anwendungsbereich konzipiert wurden. Das Prinzip der Unterscheidung von Individuen mit Hilfe einer Reihe einzelner Testaufgaben erfährt dort seine Legitimation durch die den Aufgaben unterstellte prognostische Relevanz für feststehende Tätigkeitskomplexe. Die graduelle Beherrschung dieser Tätigkeiten muß bereits als existierende Fähigkeit des Bewerbers um einen Arbeitsplatz vorausgesetzt werden.

Im selektionsbezogenen Ausbildungssystem hingegen sind Entwicklungs- und Lernkompetenzen diagnostisch zu erfassen, deren sukzessive Entfaltung und inhaltliche Strukturierung durch das System des organisierten Lernens zumindest mitgeleistet werden muß. Durch die Anwendung des formalen Prinzips der Unterscheidung von Individuen mittels objektivierter Verfahren zur Diagnose von Kompetenzen bei der Zulassung zu weiterführenden Ausbildungszweigen wird somit der Leistungsanteil des Systems in entscheidendem Maße vorweggenommen.

Der Umstand, daß im selektionsbezogenen Ausbildungssystem auch ausgewählt werden muß, kann nicht über die Unterschiedlichkeit des zu diagnostizierenden Gegenstandes hinwegtäuschen. In dem Gegenstandsbereich, für den man objektivierte Verfahren entwickelte, und die sich dort partiell bewährten (Militär, Wirtschaft, Verwaltung), ging es um das Erfassen der Beherrschung von Fähigkeiten und Fertigkeiten, die in irgendeiner Form mit den Qualifikationsanforderungen für den Arbeitsplatz zusammenhängen. Ihre Entfaltungsmöglichkeit, soweit sie überhaupt noch von Interesse bzw. durch organisierte Lernprozesse bestimmt war, fand in den materiellen Bedingungen des Arbeitsplatzes und den dort auszuführenden Tätigkeiten ihre Grenzen.

Für den diagnostischen Gegenstandsbereich der Entwicklungs- und Lernkompetenzen ist eine solche Grenzziehung nicht möglich, da

die potentielle Entfaltung dieser Kompetenzen wegen ihrer Abhängigkeit von der Qualifizierungsleistung des Ausbildungssystems nicht vorhersagbar zu konkretisieren ist. Das heißt: In der diagnostischen Erfassung von Entwicklungs- und Lernkompetenzen wird sowohl der aktuelle Lernerfolg des Schülers als auch die erwartete, aber nicht explizierte Qualifizierungsleistung des Ausbildungssystems beurteilt bzw. zu beurteilen versucht. Mit der Übertragung der Entsubjektivierungsforderung auf den diagnostischen Prozeß in der Schule werden damit alle Personen vom Beurteilungsvorgang getrennt, die noch am meisten in der Lage wären, die Qualifizierungsleistung des Ausbildungssystems einzuschätzen oder zu untersuchen, nämlich die beurteilenden Lehrer. In dieser Trennung zeigt sich der paradigmatische Charakter der Entsubjektivierungsforderung, indem objektivierte Verfahren gegenstandsunbezogen von den ursprünglichen Anwendungsbereichen auf den Gegenstandsbereich der Entwicklungs- und Lernkompetenzen übertragen werden.

Im zweiten Fall erhielten die Bemühungen um die Entsubjektivierung des diagnostischen Prozesses in der Schule ihren Anstoß durch die objektivierte Rekonstruktion der Kritik am Subjektivismus der Zensurengebung. Der Überzeugungseffekt, der von der entsubjektivierten Kritik ausgehen konnte, nahm jedoch bald paradigmatische Züge an. Die formale Darstellungsebene eines verallgemeinerbaren Dilemmas ("Subjektivismus") transformierte sich in ein formales Prinzip konstruktivistischer Forschungstätigkeit. So sehr es auch richtig ist, daß in einen diagnostischen Prozeß subjektivistische Komponenten eingehen, weil die Beurteilung komplexen Verhaltens auf eine Dimension (z.B. Notenskala) reduziert werden muß, so falsch ist die daraus gezogene Schlußfolgerung, den diagnostischen Prozeß insgesamt zu entsubjektivieren; das Reduktionismusproblem existiert nachwievor und äußert sich in der subjektivistischen Interpretation der "objektiven" Beurteilungsdaten.

Mit der Realisierung der Entsubjektivierungsforderung im konkreten, diagnostischen Prozeß ist daher nichts anderes erreicht

als die verbindliche Reglementierung des Prozesses selbst, dessen Resultate dann nachträglich auf ihre empirische Bedeutung zu untersuchen sind. Auch hier manifestiert sich die Verselbständigung der Entsubjektivierungsforderung zu einem formalen Forschungsparadigma, das als Aufforderung zur Verallgemeinerung subjektivistischer Kritik durchaus heuristisch gerechtfertigt ist, aber als konstruktivistischer Ersatz für eine Gegenstandsanalyse nur Legitimationsbedürfnisse befriedigt.

Nach der Betrachtung der bisher genannten Bedingungen für die Entsubjektivierung des diagnostischen Prozesses in der Schule läßt sich das zentrale Kennzeichen der Wirksamkeit von methodischen Paradigmen in der Verwissenschaftlichung der Schülerbeurteilung allgemein charakterisieren. Es äußert sich in der gegenstandsunbezogenen Übertragung von in einem bestimmten Gegenstandsbereich entwickelten, methodischen Vorgehensweisen der wissenschaftlichen Analyse auf einen Anwendungsbereich, in dem sich die Adäquanz des Vorgehens erst als "erkenntnisfunktional" erweisen muß. Aus der formalen Übertragung resultieren zwei Interpretationsalternativen für die Aussagen über den Untersuchungsgegenstand: Haben die Aussagen über den neuen Anwendungsbereich, für die die übertragenen methodischen Vorgehensweisen konstitutiv sind, objektive Bedeutung oder stellen sie nur methodische Artefakte dar? Die Entscheidung für eine der beiden Interpretationen läßt sich rational nicht begründen.

Typisches Beispiel dafür ist die unbegründete Annahme der Normalverteiltheit anthropologischer Merkmale, die in der schulischen Diagnostik noch häufig zur Anwendung kommt. Sie scheint durch die nicht hinterfragte Erfahrung gestützt zu werden, daß bestimmte, eindimensional erfaßte, biologische Merkmale normalverteilt sind. Das Paradigma der Normalverteiltheit hat sich bisweilen soweit verselbständigt, daß erfaßte Merkmale nur dann als untersuchungsrelevant zugelassen werden, wenn sie normalverteilt sind oder normalisiert werden können. Als Teilaspekt des Entsubjektivierungsparadigmas erfüllt das Normalverteilungsparadigma im Forschungsprozeß deshalb eine definitorische Funktion, die

durch seine Entdeckung im ursprünglichen Kontext nicht gerechtfertigt ist. In der anthropologischen Forschung läßt sich die Annahme der Normalverteilung in bestimmten Grenzen wegen der bestehenden Definiertheit der Merkmale als metrische Größen und der daraus folgenden Widerlegbarkeit der Annahme aufrechterhalten. In der Diagnostik kann man mit ihr auch aus heuristischen Gründen zumindest solange nicht arbeiten, wie die methodische Normalisierung zugelassen und damit die Hypothese der Normalverteiltheit weder zu bestätigen noch zu widerlegen ist. Dies ist natürlich nur ein problematischer Aspekt der Normalverteilung bei ihrer Verwendung in der Diagnostik. Er ist aber für die Exemplifizierung der Wirksamkeit von Paradigmen in der Forschung besonders interessant, weil er auf eine bekannte und typische Rechtfertigung der Normalverteilungsannahme verweist und das Grunddilemma sozialwissenschaftlich-empirischer Forschung verdeutlicht, mit undefinierten Größen methodisch sauber umzugehen.

"Erfahrungsgemäß zeigt sich, daß Testrohpunktwerte meist nicht normal verteilt sind. Man kann daher nicht davon ausgehen, daß die Leistungsunterschiede von Punktwert zu Punktwert gleich sind. Hier hilft folgende Annahme: An sich verteilen sich die erfaßten Leistungen normal. Wenn bei den empirischen Rohpunktwerten nun nicht-normale Verteilungen auftreten, so liegt dies in den Eigenarten der Aufgaben begründet, welche die Rohpunktskala verzerren. Wenn es nun gelingt, diese Verzerrung rückgängig zu machen, so ergibt sich wieder eine Normalverteilung. Dieses Rückgängigmachen der Verzerrung erreicht man durch die Transformation der Rohpunkte ..." (LANGFELDT, 1974, S. 119).

Wie man leicht erkennen kann, rechtfertigt sich bei LANGFELDT die Normalverteilungsannahme durch eine tautologische Argumentation: Unter der Voraussetzung, daß eine bestimmte Verteilungsform von Daten in eine Normalverteilung entzerrt werden kann, läßt sich die Annahme aufrechterhalten, daß die "wahre" Normalverteilung durch die durch die Daten gegebene Verteilungsform verzerrt wurde.

Diese Rechtfertigungsprozedur ist bedauerlicherweise ein typisches Verfahren in der empiristisch orientierten, sozialwissenschaftlichen Analyse von Gegenstandsbereichen. Wird ein Komplex von Methoden der wissenschaftlichen Analyse auf tautologischem

Weg in die Forschung eingeführt oder setzt sich ein solcher in der konkreten, wissenschaftlichen Tätigkeit allmählich durch, ohne in seiner tautologischen Funktion erkannt zu werden, so läßt er sich für die Untersuchung beliebig vieler Gegenstandsbereiche heranziehen. Dabei kann weder die Adäquanz der Methoden für den jeweiligen Gegenstand hinterfragt werden, noch ist die empirische Bedeutung der Untersuchungsergebnisse anzugeben. So ist z.B. die Annahme des "wahren Wertes" in der klassischen Testtheorie eine methodenbedingte Notwendigkeit, deren meßtheoretische Voraussetzungen in der Diagnostik nicht erfüllt sind, sodaß die Deutbarkeit des "wahren Wertes" ins Belieben des einzelnen Forschers gestellt bleibt.

"Man geht von der Vorstellung aus, daß die Testung einer Person mit einem Test 'fehlerhaft' ist und daß daher jeder empirischen Messung ein 'w a h r e r W e r t' entspricht, welcher sich von dem beobachteten durch eben jene Meßfehler unterscheidet. Handelt es sich etwa um einen Intelligenztest, dann wäre der 'true score' die 'wahre Intelligenzleistungsfähigkeit' der betreffenden Person in dem verwendeten Test zum Zeitpunkt der Testvorgabe. Was Intelligenz ist, wird durch den verwendeten Test definiert, d.h. Intelligenz ist, was der Test mißt. Daß eine derartige Testtheorie, auch was die psychologische Theorienbildung anbelangt, in Schwierigkeiten führen muß, ergibt sich daraus, daß jeder Test eine neue Definition von 'Intelligenz' beinhaltet (FISCHER, 1974, S. 27).

1.3.2.4 Das Forschungsparadigma der Entsubjektivierung als Ausdruck der Technologisierung gesellschaftlicher Prozesse

Die tautologische Rechtfertigung von Methoden in der Wissenschaft ist lediglich die Konsequenz der gegenstandsunbezogenen Übernahme dieser Methoden von ihrem ursprünglichen Konstitutionsbereich für einen neuen Anwendungsbereich. In ihrer unreflektierten Verwendung in diesem neuen Bereich äußert sich ihre paradigmatische Funktion für die konkrete wissenschaftliche Tätigkeit. In der Beliebigkeit der Interpretation der mit ihrer Anwendung gewonnenen Ergebnisse zeigt sich zwar deutlich, daß sie keinerlei Erkenntnisgewinn erbringen können; dies ist aber unter

bestimmten pragmatischen, gesellschaftlichen Bedingungen ihr Vorteil. Wenn mit der Übernahme solcher Methoden eine Objektivierung bereichsspezifischer Forschungstätigkeit verbunden ist, die gesellschaftlich durchgesetzten Bestrebungen der Rationalisierung und Technologisierung entgegankommt, dann erweist sich die Interpretationsbeliebigkeit der Ergebnisse methodisch objektivierter Forschung sogar als besonders nützlich. Dann bedarf selbst die inhaltliche Orientierung der wissenschaftlichen Tätigkeit keiner Legitimation mehr.

Gelingt es, die Untersuchungstätigkeit von Wissenschaftlern durch den Filter der methodischen Objektivierung zu verallgemeinern, so tragen die Ergebnisse der Untersuchungstätigkeit den Schein der Allgemeingültigkeit, auch wenn sich deren Konstitutiertheit im gemeinsamen methodischen Vorgehen der Wissenschaftler erschöpft. Für den zeitlichen Anfang der Verwissenschaftlichung der Diagnostik in der Schule ist deshalb zu untersuchen, in welcher Form die Allgemeingültigkeit von Urteilsaussagen über Schüler unter Berücksichtigung der Erfahrungen mit objektivierten Beurteilungsprozessen und der objektivierten Kritik am Lehrerurteil erscheinen konnte und von welchen gesellschaftlichen Bedingungen die Reformbestrebungen abhingen. Hier sind die entscheidenden Ursachen festzumachen, die die sukzessive Durchsetzung des Entsubjektivierungsparadigmas in der schulischen Diagnostik bedingten.

In der Zeit nach dem ersten Weltkrieg, in der in Deutschland mit der Entwicklung von Schultests begonnen wurde, vollzogen sich die verschiedenen Veränderungen des gesamten Schulwesens im Spannungsfeld zwischen reaktionären und reformerischen Elementen. Den politischen und ökonomischen Hintergrund dieser Veränderungen bilden die sich in Industrie und Wirtschaft durchsetzenden Rationalisierungs- und Technologisierungstendenzen, deren Voraussetzungen sich zwar bereits seit der Industrialisierung im 19. Jahrhundert entwickelten, in der Aufbauphase nach dem ersten Weltkrieg jedoch zur vollen Entfaltung gelangten. In seiner Betrachtung des Verhältnisses von Kapitalentwicklung

und Qualifikationsentwicklung weist OTTEN darauf hin, daß vor diesem Zeitpunkt die entscheidende Produktivitätssteigerung, d.h. die Veränderung der Arbeitsplatzstrukturen, eher eine arbeitsorganisatorische als eine technologische Innovation war (1973, S. 69). Mit der technologisch-strukturellen Veränderung der Produktionsmittel vollzog sich eine vollständige Umstrukturierung des Arbeitsprozesses und der damit verbundenen Qualifikationsanforderungen. Die ihr vorangegangene, erste Stufe der Rationalisierung und Technologisierung in Form der Fließbandarbeit, die sich vor allem in den USA relativ früh durchsetzte, führte im Prinzip eher insgesamt zur Dequalifikation des Arbeitsvermögens. Die zweite Stufe der Entwicklung durch die beginnende Verwissenschaftlichung des Produktionsprozesses in Form zunehmender Automation erforderte eine partielle Höherqualifizierung der Arbeitskraft bei gleichzeitiger Hierarchisierung der Qualifikationsstruktur des Gesamtarbeiters. Ebenso erfuhren die Beschäftigungsbereiche der privaten und staatlichen Dienstleistung sowie die kaufmännischen Berufszweige eine explosive Ausweitung. Die Entwicklungstendenz, die hier nur angedeutet werden kann, konkretisierte sich in der wachsenden Zunahme der Beschäftigten in den verschiedenen, strukturell neu gegliederten Produktionsbereichen und in der permanenten Ausweitung der Rekrutierungsbasis, z.B. der Beschäftigung von Frauen im Büro (vgl. BAHRDT, 1958).

Die Konsequenzen dieser Entwicklung führen im gleichen Ausmaß zur Problematisierung der Ausbildungsfunktion des Bildungswesens als Qualifizierungsinstitution für die Arbeitskraft. Für die veränderte Aufgabenstellung des Ausbildungssystems kann damals schon im Ansatz gelten, was WINNACKER, der Vorsitzende des Vorstandes der Farbwerke Hoechst AG, für die Qualifizierungsfunktion des Bildungswesens später forderte:

"..., daß eine wachsende Schicht von Technikern, Facharbeitern und Ingenieuren mit wachsenden Qualifikationen für die Monopole von zunehmenden Interesse ist und daß dies wiederum von einer verbesserten wissenschaftlichen Ausbildung der Ausbilder und der Spitzeningenieure abhängt. Allein aus diesem Grund, d.h. vorwiegend zum Zweck der optimalen Rekrutierung bedarf es einer

fundamentalen Reform der Ausbildung, die ihren Ausgang im Schulwesen nehmen muß. Daher fällt dem Staat die Aufgabe zu, zweierlei zu garantieren, erstens eine breit angelegte Bildung und zweitens stets modifizierbare Qualifikationen, um diese an die sich immer schneller wandelnden Ansprüche des Kapitals anpassen zu können" (zit. n. OTTEN, 1973, S. 72).

Im Rahmen des Versuches, den Einfluß gesellschaftlicher Bedingungen auf das Wirksamwerden des Entsubjektivierungsparadigmas zu bestimmen, kommt es darauf an abzuschätzen, inwieweit die Objektivierungstendenzen in der Schülerbeurteilung notwendige Erscheinungsform der Bewertung von Qualifikationen sind, deren Produktion durch das Schulsystem vor dem Hintergrund der ökonomischen Entwicklung in der Aufbauphase nach dem ersten Weltkrieg in Deutschland erforderlich wird. Die Veränderung der Produktionsweise zu diesem Zeitpunkt ist durch den Beginn einer enormen Beeinflussung durch die Wissenschaft gekennzeichnet. Mit der Verwissenschaftlichung der Produktion, d.h. mit der technologischen Umsetzung von Erkenntnissen der Wissenschaft - insbesondere der Naturwissenschaft - wird eine veränderte arbeitsteilige Strukturierung des Arbeitsvollzuges im Produktionsprozeß möglich. Sie hängt in starkem Maße von den wissenschaftlich untersuchten Gesetzmäßigkeiten in der Beherrschung der Natur bzw. von den Zusammenhängen zwischen solchen Gesetzmäßigkeiten ab und sie erlaubt die technische Instrumentalisierung der Situationen im Forschungslabor in großem Umfang und auf Massengrundbasis.

Charakteristisch für solche Arbeitsvollzüge ist ein formales Merkmal naturwissenschaftlicher Forschung: die Realisierung eines materiellen Prozesses unter kontrollierten, konstant gehaltenen Bedingungen. Durch die Konstanzhaltung bestimmter Bedingungen lassen sich materielle Prozesse wiederholen bzw. als wiederholbar bezeichnen und somit verallgemeinern und vermitteln, d.h. in der Form eines gesetzesartigen Zusammenhanges ausdrücken. Während aber im naturwissenschaftlichen Forschungsprozeß die Konstanzhaltung der Bedingungen durch das forschende Subjekt hergestellt und kontrolliert wird, um die Wiederholbarkeit

materieller Prozesse konstatieren zu können, liegen die Verhältnisse im konkreten Arbeitsprozeß genau umgekehrt: Der materielle Prozeß und die invarianten Bedingungen für die Wiederholbarkeit seines Ablaufes sind dem Arbeitenden als Technologie vorgegeben und der Arbeitende hat die Aufgabe, den Prozeß und seine Wiederholung im Rahmen der durch die Arbeitssituation zugelassenen Grenzen zu realisieren.

Soweit der Arbeitsvollzug durch die technologisierten Naturgesetzmäßigkeiten determiniert ist, muß der Arbeitende eine Menge oder einen Komplex von repetitiven Operationen ausführen, deren Zusammenhang durch das erzielte Arbeitsprodukt festgelegt ist. Das, was also an der Tätigkeit des Arbeitenden als wiederholbar bzw. verallgemeinerbar festgehalten und in objektivierter Form ausgedrückt werden kann, ist durch die technische und technologische Struktur des Arbeitsvollzuges determiniert.⁴⁾

-
- 4) Der Vergleich des naturwissenschaftlichen Forschungsprozesses mit dem Arbeitsvollzug in der Produktion ist sicherlich stark vereinfacht und scheint sich nur auf die Situation eines Fließbandarbeiters zu beziehen. Ich halte aber die Generalisierung für legitim, um den Grundgedanken einer subjektivierten Qualifikationsbewertung herausarbeiten zu können. Überdies erscheint mir die Generalisierung auch dann zulässig, wenn man in einer detaillierten Analyse das Problem der Determiniertheit des Arbeitsvollzuges durch die technologisierten Naturgesetzmäßigkeiten untersuchen wollte und deren Funktion für die Bestimmung von Qualifikationen. Es würde sich dabei zeigen, daß relativ exakte Qualifikationsbestimmungen nur bei starker Determiniertheit möglich sind, andernfalls aber subjektivistische Komponenten an Einfluß gewinnen. Deutlich tritt dieses Problem bei der Qualifikationsbestimmung für Führungskräfte in Erscheinung. Der Globalitätsgrad ist hier so hoch, daß damit für den konkreten Beurteilungsprozeß nichts anzufangen ist. Deshalb wurde für die Beurteilung des Topmanagements die Konsequenz in der Weise daraus gezogen, sich mehr an der privaten Propaganda und an der Unternehmensbilanz zu orientieren. Völlig unberücksichtigt bleibt hier die Betrachtung der historischen Entwicklung der Verwissenschaftlichung gesellschaftlicher Produktion, sowie des Einflusses der Produktionsweise auf die Möglichkeit der Entstehung und Verselbständigung objektivierter Naturerkenntnis. Erst durch eine solche Untersuchung ließe sich das angedeutete Problem des Zusammenhanges von Arbeit und Qualifikation in der notwendigen Weise präzisieren. Erste erfolgversprechende Ansätze wurden dazu vom

Selbstverständlich kann das Verfahren der technologischen Strukturierung des Arbeitsvollzuges auch auf Arbeitsbereiche übertragen werden, in denen keine naturwissenschaftlichen Gesetzmäßigkeiten die Grundlage für die am Arbeitsplatz auszuführenden Tätigkeiten darstellen. Die Rationalisierung und Technologisierung in den kaufmännischen Beschäftigungsbereichen sei hier nur als Beispiel angeführt. Insgesamt ist für die Durchsetzung des Entsubjektivierungsparadigmas in der empiristisch orientierten, sozialwissenschaftlichen Forschung die Tatsache wichtig, daß in der rationalisierten und technologisierten Produktionsstruktur und den darin erforderlichen Arbeitsvollzügen die Voraussetzung für die Wiederholbarkeit bestimmter Tätigkeiten bzw. Tätigkeitskomplexe geschaffen wird. Nur die Wiederholbarkeit und damit Verallgemeinerbarkeit bildet die objektive Bedingung für die Objektivierung des Arbeitsverhaltens in Form von Qualifikationen, Fähigkeiten, Fertigkeiten usw. Mit der Notwendigkeit Arbeitsvollzüge auszuführen, deren Struktur und Ablauf durch die Technologie und Maschinerie geprägt sind, und mit der Aufstellbarkeit der Forderung nach menschlichen Kompetenzen, die zu einer wiederholten Ausführung solcher Arbeitsvollzüge befähigen, läßt sich die Konstruktion von Qualifikationsmustern der Form nach rechtfertigen. Sie fassen im Querschnitt wichtige, allgemeine Tätigkeiten aus verschiedenen Beschäftigungsbereichen zusammen und ihre Beherrschung kann zur Voraussetzung für die Zulassung zu einem Arbeitsplatz gemacht werden.

Der konkrete Konstitutionsprozeß der jeweils historisch durchgesetzten Qualifikationsmuster ist wahrscheinlich nur noch in den seltensten Fällen rekonstruierbar (vgl. z.B. die Untersu-

AUTORENKOLLEKTIV AM INSTITUT FÜR SOZIOLOGIE DER FU BERLIN (1973) und von ENGELHARDT & HOFFMANN (1974) geliefert, in denen auch die Frage der Vergesellschaftung des Wissenschaftsprozesses angegangen wird. Letzteres ist insbesondere für die z.Z. beginnende, gesellschaftskritische Reflexion naturwissenschaftlicher Forschungstätigkeit von Bedeutung (vgl. DAS ARGUMENT 88, 1974), weil hier die Untersuchung der Entwicklung des Entsubjektivierungsparadigmas in seiner gesellschaftlichen Determiniertheit anzusetzen hat.

chungen zum Einfluß der Unternehmer auf die Bildungspolitik in der BRD von BAETHGE, 1970). Es läßt sich aber allgemein behaupten, daß Qualifikationsmuster in dem Maße eine entsubjektivier- te Bestimmungsgrundlage haben, als darin die Verallgemeinerung des in der Technologie und Maschinerie bestimmter Beschäftigungs- bereiche vergegenständlichten Wissens und der vergegenständlich- ten Fähigkeiten und Fertigkeiten eingegangen ist.

In der technologiebedingten Objektivierungsmöglichkeit solcher auf die Produktion beziehbarer Qualifikationsmuster muß der An- satz und die Voraussetzung für die Schulreformtendenzen zu Be- ginn der Weimarer Republik gesehen werden. Denn "die Ausbil- dung von technischer Arbeitskraft gemäß den Anforderungen der als Technologie und Maschinerie vergegenständlichten Arbeit be- deutet ... die Erstellung von allgemeinen Qualifikationen, die dazu befähigen, die Funktion der Handhabung und Kontrolle be- reits vergegenständlichten Wissens wahrzunehmen, und zwar auf verschiedenen Niveaus der technischen Qualifikationen" (AUTO- RENKOLLEKTIV, 1973, S. 194). Diese Qualifikationsmuster wurden zwar nicht explizit und verallgemeinert definiert bzw. als Ausbildungsziele vorgeschrieben, sie setzten sich aber im Zuge der zunehmenden Vergesellschaftung der Produktion allmählich und branchenspezifisch unterschiedlich verbindlich durch und waren von subjektivistischen Momenten und offiziell existieren- den Ideologien überlagert.

Die Erstellung von allgemeinen Qualifikationen konkretisierte sich damals in der institutionelle Verankerung von verschiede- nen Reformtendenzen, in denen die Beziehung der Ausbildung zu den gesellschaftlich durchgesetzten und produktionsrelevanten Qualifikationsanforderungen noch sehr global vermittelt war und lehrinhaltlich nur wenig zum Ausdruck kam. Sie äußert sich im sukzessiven Ausbau der Realschulen, die ursprünglich als eine Mittelstandsschule für das selbständige Gewerbe konzipiert wa- ren, aber dann die Rolle einer Qualifikationsinstanz für eine gehobene Art von kommerziellen Lohnarbeitern übernahmen (OTTEN, 1974). Die Grundschulreform von 1920, an die anfänglich auch der Gedanke des Abbaus der Chancenungleichheit geknüpft war, führte

zu einer allgemein verbindlichen Ausbildung in den untersten vier Jahrgängen zur Beherrschung der einfachsten Kulturtechniken des Lesens, Schreibens und Rechnens mit dem gleichzeitigen Anspruch, daß sie "den Bestbeanlagten das Maß von Förderungen zuteil werden läßt, das den Kräften dieser Kinder angemessen ist" (DEUTSCHER STÄDTETAG, 1925, S. 229).

Drückt sich in der Forderung nach der allgemein verbindlichen Vermittlung der einfachsten Kulturtechniken der breiteste Querschnitt notwendig zu erwerbender Qualifikationen aus, so steht die Zielsetzung, die Bestbeanlagten besonders zu fördern, für die stärker hervortretende Erkenntnis, daß der Schule auch eine Vorbereitungsfunktion für die Ausbildung einer technisb-h-wissenschaftlichen Elite zukommen muß, um den zunehmenden, industriellen Bedarf unter dem wachsenden interantionalen Konkurrenzdruck zu decken. Dies wird auch bestätigt durch die Artikulation des Bedürfnisses nach höheren Fachschulen und den Ausbau des Ingenieurschulwesens (RIEDLER, 1916). In der Entwicklung des Berufsschulwesens war eine starke Produktionsbezogenheit der Ausbildung allein schon durch ihre dezentralisierte Struktur gegeben. Die staatliche Überwachung dieser Ausrichtung erfolgte durch das Ministerium für Handel und Gewerbe, das die inhaltlichen Richtlinien festlegte, denen zufolge die Vertiefung der deutschen Sprachkenntnisse und die Aneignung einiger Rechenoperationen, die für die Warenkunde relevant sind, von besonderer Bedeutung waren (OTTEN, 1974, S. 81f).

Unabhängig von noch differenzierteren Veränderungen des Ausbildungssystems während der 20er und 30er Jahre läßt sich das Ergebnis der schulischen Reformbewegung durch die Tendenz typisieren, die institutionelle und strukturelle Untergliederung des Ausbildungswesens an allgemeinen Qualifikationsmustern zu orientieren, die ihre objektiven, gesellschaftlichen Voraussetzungen in den Arbeitsvollzügen in bestimmten verallgemeinerten Produktionsbereichen haben. Mit dieser durch produktionsbezogene Qualifikationsbedürfnisse geprägten Vermitteltheit von Bildungswesen und Gesellschaft war gleichzeitig der Boden für die formale

Übernahme der in der Produktion wirksamen Grundmodelle und -vorstellungen der Qualifikationsbewertung vorbereitet.

Systematisch betrachtet, hat das Grundmodell der Bewertung von Qualifikationen seinen Ursprung in der nach technologischen Prinzipien organisierten Struktur des Produktionsablaufes, in der arbeitsteilig gegliederte und mehr oder weniger komplexe Arbeitsvollzüge innerhalb der wiederum technologisch gegebenen Realisierungsgrenzen wiederholt durchgeführt werden müssen. Die dadurch bedingte Objektivierung der konkreten Arbeitstätigkeit als Funktion der Technologie, d.h. als standardisierbare Tätigkeit im erforderlichen Umgang mit der Technologie, erlaubt es, die standardisierbaren Arbeitsvollzüge in der Zeit zu bewerten, als Grad der Beherrschung auszugeben und je nach den materiell-qualitativen Eigenschaften des Arbeitsprozesses und der Verwendung des Arbeitsproduktes als eine bestimmte Qualifikation zu bezeichnen. Die Struktur des Grundmodells der Bewertung von Qualifikationen tritt in hervorragender Weise in den wissenschaftlichen Arbeitsanalysen (GHISELLI & BROWN, 1955) sowie in den Rationalisierungsversuchen des Arbeitsverlaufes (REFA) zutage, wo die Objektivierung der Tätigkeit in zeitoptimalen Arbeitsperioden erzwungen wird. Dieses Grundmodell ist auch Bestandteil einer modifizierten Form der Qualifikationsbewertung, deren formale Kennzeichnung sich in der Testkonstruktion wiederfindet: Bei Arbeitsvollzügen, die sich in mehrere Teilvollzüge technologisch untergliedern lassen, aber insgesamt von einer Person ausgeführt werden, kann man die Teilvollzüge getrennt bewerten und die Einzelbewertungen nachträglich additiv oder in einer anderen Weise linear kombinieren.

Damit ist mit der technologiebedingten Objektivierung des Arbeitsvollzuges die materielle Basis für die Entsubjektivierung der Bewertung der dadurch festgelegten Qualifiktionen geschaffen und die Möglichkeit gegeben, den zu bewertenden, standardisierbaren Arbeitsvollzug durch diagnostische Situationen für prognostische Zwecke in methodisch objektivierter Form zu beurteilen. Die diagnostischen Situationen müssen dabei so instrumentalisieren

sein, daß sie die später auszuführenden, verallgemeinerbaren Arbeitsvollzüge in gewisser Weise vorwegnehmen, sie repräsentieren bzw. ein Modell derselben darstellen. In der Repräsentativitätsforderung an die diagnostischen Situationen liegen letztlich die theoretischen Schwierigkeiten für eine prognostisch relevante Interpretation der Ergebnisse objektivierter Beurteilungen begründet. Die Auswahl und Kombination der diagnostischen Modellsituationen bzw. der Reaktionen der Versuchspersonen auf sie, läßt sich durch keinen theoretischen Bezugsrahmen im Sinne eines Ableitungszusammenhanges innerhalb einer Qualifikationstheorie synthetisieren. Weder deduktive noch induktive Ableitungsmodelle bieten hier eine wissenschaftlich fundierbare Hilfestellung (MEYER, 1972).

Der realisierte Lösungsausweg über das testtheoretische Diskriminanzprinzip und der empiristischen Validitätsüberprüfung hat aber den bereits weiter oben erwähnten Vorteil: Weil die theoretische Problemstellung nicht bewältigt werden kann, lassen sich unter das Diskriminanzprinzip beliebige, inhaltlich unterschiedliche, diagnostische Fragestellungen subsumieren, d.h., die objektivierte Beurteilung einer technologiebezogenen Qualifikation ist dann diagnostisch signifikant und im testtheoretischen Sinne prognostisch relevant, wenn sie innerhalb der Anwendungspopulation in gleicher Weise diskriminiert wie das zur Validitätsüberprüfung des Beurteilungsverfahrens herangezogene Kriterium, z.B. eine langfristige Bewährung am Arbeitsplatz. Durch die geltungspragmatische Prioritierung des Diskriminanzprinzips bei der Auswahl und Zusammenstellung von diagnostischen Situationen wird somit die Inhaltlichkeit der Qualifikationsdimension gegenüber der methodischen Form ihrer Objektivierung austauschbar.

Dies ist ein entscheidendes Merkmal objektivierter Beurteilungsverfahren für ihre Übertragung in den schulischen Bereich. Daß neben den in der Repräsentationsforderung steckenden Schwierigkeiten auch andere Einwände gegen die formale Übertragung des Modells kritisch zu überprüfen sind - insbesondere jene, die

sich auf die unterstellte Annahme der Konsistenz des Lernverhaltens in der Zeit erstrecken -, sei vorerst nur erwähnt. Wichtig ist hier, daß mit der, wenn auch nur, kategorialen Beziehung der Ausbildungsreform während der 20er und 30er Jahre zu den produktionsbezogenen Qualifizierungsbedürfnissen der Grundstein für das Platzgreifen von Modellen der Qualifikationsbewertung in der Produktion im schulischen Bereich gelegt wurde und daß erst dadurch das gesellschaftlich durchgesetzte Entsubjektivierungsparadigma im Rahmen der Beurteilung von Dispositionen zur Orientierungslinie für die Entwicklungstendenzen im Bereich der Diagnostik in der Schule werden konnte.

Vor diesem Hintergrund wird erst verständlich, warum die Erfolgserwartungen bei der Entwicklung und Anwendung von objektivierten Verfahren in außerschulischen Bereichen auch in der Schule als richtungsweisend fungierten und warum die bekannte Kritik am Subjektivismus des Lehrerurteils einer Objektivierung unterzogen wurde bzw. ihre Verkehrung in die Entsubjektivierung des gesamten diagnostischen Prozesses bei der Schülerbeurteilung anzusetzen begann. All dies wurde erst ermöglicht durch die sukzessive Entwicklung eines Objektivierungsprozesses, der sich ungeachtet der qualitativen Beschaffenheit des zu untersuchenden Gegenstandes durchführen ließ; für ihn war die Eigenschaft der Diskriminierungsfähigkeit auf formaler Ebene konstitutiv, was sowohl im Einklang mit der Grundvorstellung der Qualifikationsbewertung in der Produktion als auch mit dem Selektionsprinzip des Ausbildungswesens stand.

An dieser Grundvorstellung hat sich bis heute ebensowenig geändert wie an der Wirksamkeit des Selektionsprinzips. Auch die aktuellen Versuche zur Verwissenschaftlichung der Schülerbeurteilung finden deshalb in der Berücksichtigung dieser beiden Bestimmungsmomente ihre Grenzen. Darin dürfte das Fehlen einer eigenständigen, p ä d a g o g i s c h e n Diagnostik zu begründen sein. Darüberhinaus wurden die Versuche zur Entsubjektivierung der Schülerbeurteilung unterstützt durch die sukzessive Orientierung von Sozialwissenschaftlern an der Entwick-

lung der empiristischen Wissenschaftstheorie, die als systematisierte Reflexion naturwissenschaftlicher Forschung - in erster Linie der exakt-naturwissenschaftlichen Forschung - die Begründung der Entsubjektivierung wissenschaftlicher Aussagen zum Ziel hatte. Ob die Orientierung der diagnostischen Forschung an der empiristischen Wissenschaftstheorie konsequent war und ob sie der Analyse des diagnostischen Gegenstandsbereiches gerecht werden kann, steht im Mittelpunkt der folgenden Betrachtung. Es wird sich dabei zeigen, daß die Zuhilfenahme von wissenschaftstheoretischen Modellen zur Begründung von allgemein gültigen Aussagen in der Diagnostik das Vorgehen in der Verwissenschaftlichung der Schülerbeurteilung nur legitimieren soll und die Entsubjektivierungsversuche nur forschungsparadigmatische Bedeutung haben.

1.4 Der Stellenwert von Forschungsparadigmen im Wissenschafts-prozeß

Bei dem Versuch, die empiristisch orientierte Forschung im Bereich der Diagnostik in der Schule durch die sukzessive Anlehnung an das Entsubjektivierungsparadigma zu kennzeichnen, um dadurch ihre Gegenstandsunbezagenheit herauszuarbeiten, wurde der Begriff des Forschungsparadigmas relativ unsystematisiert verwendet. In der Beschränkung auf seine exemplarische Veranschaulichung und der Untersuchung wesentlicher Bedingungen für das Wirksamwerden des Entsubjektivierungsparadigmas in der Diagnostik sollte zugleich die Zurückhaltung vorweggenommen werden, die sich mit der Einführung eines neuartigen Untersuchungsansatzes und eines in der Wissenschaftstheorie nicht unbestrittenen Begriffes verbinden muß.

Seine Funktion innerhalb dieser Arbeit besteht nicht darin, die bestehenden Forschungsansätze und vorhandenen Untersuchungsergebnisse in der verwissenschaftlichten Diagnostik in der Schule begrifflich zu synthetisieren und daraus Perspektiven für eine Neuorientierung zu konstruieren; dies kann durch theoretische Reflexionen ohne entsprechenden Praxisbezug nicht geleistet werden. Der Paradigmabegriff dient vielmehr der strukturellen Bestandsaufnahme einer weitgehend irrational verlaufenden, wissenschaftlichen Tätigkeit auf dem Gebiet der schulischen Diagnostik, in der man durch die Verwendung einer empiristischen Terminologie den Anspruch auf methodisch sauberes, wissenschaftliches Arbeiten anmeldet. In der Reflexion der Berechtigung dieser Terminologie und der Konsequenzen aus dem methodischen Vorgehen wird der gleiche Anspruch jedoch de facto preisgegeben.

Es liegt nicht in meiner Untersuchungsabsicht, die empiristische Orientierung wissenschaftlicher Analysen des diagnostischen Prozesses rundherum abzulehnen, um den Weg für die Rehabilitation des kritisierten Lehrerurteils freizumachen. Ich glaube aber, daß es angesichts der hektischen Aktivität auf dem Sektor der Testproduktion und des zunehmenden Objektivierungs-

druckes auf die Zulassungsprozeduren im Ausbildungswesens darauf ankommt, auf der reflektorischen Ebene die feststellbare Irrationalität in der Verwissenschaftlichung der Schülerbeurteilung wegen der daraus resultierenden Konsequenzen für die Beurteilten und für bildungspolitische Reformmaßnahmen aufzuzeigen.

1.4.1 "Irrationalität" in der Wissenschaft ?

Wollte man den Stellenwert des Paradigmbegriffes zum Zeitpunkt seines Aufkommens in der Wissenschaftstheorie auf einen stark verkürzten und simplifizierten Nenner bringen, so könnte man sein wesentliches Merkmal in der auch hier verwendeten Funktion zusammenfassen, nämlich in der Kennzeichnung der Irrationalität in der Wissenschaftsentwicklung. Der Begriff "irrational" impliziert allerdings ein breites Spektrum von unterschiedlichen Bedeutungen, die sich aus der Negation dessen ergeben, was jeweils mit rational gemeint ist. Der Vorwurf der simplifizierten Kennzeichnung der Wissenschaftsentwicklung bestünde daher zu recht.

Eine sinnvolle Eingrenzung des Bedeutungsspektrums wird aber durch die Tatsache ermöglicht, daß sich der geistige Vater des Paradigmakonzeptes mit dem Versuch der wissenschaftsgeschichtlichen Rekonstruktion der Theorienentstehung und -entwicklung gegen die Rationalitätsvorstellungen der logisch-empiristischen Wissenschaftstheorie richtete. Als KUHN im Jahre 1962 sein Buch "The structure of scientific revolutions" veröffentlichte, begab er sich mit seinen Thesen in das Feld einer wissenschaftstheoretischen Auseinandersetzung, in der man sich um die Kennzeichnung einer zeitlosen, empirischen Basis oder empiristischen Sprache bemühte, um die Aufstellung eines allgemeinen Sinn- oder Abgrenzungskriteriums wissenschaftlicher Aussagen, um die Bestimmung des Verhältnisses von Theorie und Erfahrung, um die Entwicklung eines formalen Modelles wissenschaftlicher Erklärung

und schließlich um die logische Behandlung des Induktionsproblems. Bestimmte Fragestellungen waren bereits früher und zum Teil erfolglos behandelt worden. Insgesamt ist aber für die seinerzeitige Diskussion der Probleme der ahistorische Standpunkt typisch.

Dementsprechend mußte das wissenschaftsgeschichtliche Vorgehen von KUHN gegen die REICHENBACHsche Unterscheidung von Entstehungs- und Begründungszusammenhängen für wissenschaftliche Theorienbildung und -entwicklung von den empiristischen Wissenschaftstheoretikern als Herausforderung begriffen werden. KUHN zeichnet ein völlig verändertes Bild der Wissenschaftsentwicklung:

"die Wissenschaften schreiten voran in 'normalen' Phasen der Ausgestaltung und Anwendung paradigmatischer Theorien und Methoden, die einander in krisenhaften Erschütterungen und revolutionären Neuorientierungen ablösen. Solche 'wissenschaftlichen Revolutionen' sind nach Kuhn weniger logisch als psychologisch und soziologisch zu kennzeichnen, da das alte vorrevolutionäre 'Paradigma' mit dem neuen 'inkommensurabel' ist; die Wissenschaftler werden nicht von der Falschheit des alten und der Richtigkeit des neuen Paradigmas überzeugt, sondern die neue revolutionäre Sicht der Dinge setzt sich nach individuellen 'Bekehrungserlebnissen' von der Art eines 'gestalt switch' bei zunächst nur wenigen, meist jungen Wissenschaftlern durch und wird Kern des wissenschaftlichen Programms der rekonstruierten Wissenschaftler-Gemeinschaft (scientific community) der betreffenden Disziplin. Die Anhänger des alten Paradigmas lassen sich, von der Fruchtbarkeit der neuen eingeschlagenen Forschungsrichtung beeindruckt, schließlich überzeugen oder zur Teilnahme gewinnen oder leisten, immer vereinzelter, nachlassenden Widerstand und sterben schließlich aus" (DIEDERICH, 1974, S. 8f).

KUHN hat seine "Wissenschaftsentwicklung" für den Bereich der exakten Naturwissenschaften konzipiert. Dies hervorzuheben, ist wichtig, weil sich daraus für die Übertragung seines Konzeptes, der Weiterentwicklung dieses Konzeptes durch ihn und des darauf bezogenen wissenschaftstheoretischen Rekonstruktionsversuches durch STEGMÜLLER bzw. SNEED eine spezifische Problematik ergibt: Es wird sich zeigen, daß daran geübte Kritiken zurückzuweisen sind, wenn sich hinter dem intendierten Anwendungsbereich die Naturwissenschaften verbergen, nicht aber, wenn man gewisse kritische Einwände gegenüber KUHN auf die Diagnostik in der Schule als Teilaspekt der empiristisch orientierten Sozialwissenschaft-

ten bezieht.

Dies ist umso bedeutsamer, als KUHN psychologische und soziologische Faktoren des Konstitutionsprozesses von Naturwissenschaften in den Vordergrund stellt und seine Gegner wie z.B. POPPER und SCHEFFLER ihm deshalb "Subjektivismus" vorwerfen. Das würde wiederum unter Bezug auf die empiristisch orientierten Sozialwissenschaften bedeuten, daß, solange deren empiristisches Vorgehen nicht den Exaktheitsforderungen der Kuhn-kritiker entspricht, der Subjektivismusvorwurf nicht in dem Sinn abgewiesen werden kann, wie dies von den wissenschaftstheoretischen Befürwortern der Kuhnschen Auffassung erfolgt, die die psychologischen und soziologischen Komponenten in eine "rationale" Rekonstruktion integrieren wollen. Bevor ich zur Diskussion dieses Problems übergehe, bedarf die vorher skizzierte, zusammenfassende Darstellung der Position KUHNs sowie die kritische Reaktion darauf einer differenzierenden Erörterung.

1.4.2 Darstellung des KUHNschen Paradigmagedankens

Interessanterweise bemüht sich KUHN in anbetracht des historisch zu verfolgenden Fortschritts in den Naturwissenschaften und des bestehenden und differenzierten Gebäudes seines wissenschaftstheoretischen Rekonstruktionsversuches als sukzessive Ausweitung und Anhäufung von Wissen um den Nachweis, daß man die Entwicklung der Einzelwissenschaften als einen vollkommen "irrationalen" Prozeß deuten muß (In Relation zu den Naturwissenschaften dürfte die empiristisch orientierte Sozialwissenschaft keine vergleichbar reflektierte Erfolgsbilanz aufzuweisen haben) ⁵⁾.

5) Ich beziehe mich zur Schilderung der KUHNschen Position auf die einschlägige Darstellung durch STEGMÜLLER (1973, Band II, Studienausgabe Teil E). Die Gründe dafür sind, daß STEGMÜLLER dem KUHNschen Standpunkt nicht nur gegenübersteht, sonder aus wissenschaftstheoretischer Perspektive den Versuch der rationalen Rekonstruktion dieser Position unternimmt. Darüberhinaus dürfte nicht zuletzt aus den genannten Gründen

Dazu bedient er sich dreier zentraler Begriffe: "normale Wissenschaft", "Paradigma" und "wissenschaftliche Revolution". Die normale Wissenschaft hat es mit dem Lösen von Rätseln zu tun, d.h., sie hat die Aufgabe, historisch bestehende, unterschiedliche Deutungen und Beschreibungen von Phänomenen eines bestimmten Phänomenbereiches auf deren Zusammenhang zu untersuchen und abweichende Vorstellungen darüber theoretisch zu synthetisieren. Die Tätigkeit "normaler Wissenschaftler" vollzieht sich allerdings immer im Kontext eines richtungsweisenden, tradierten und integrierenden Paradigmas, das für eine gewisse Zeit einer Gemeinschaft von Fachleuten Modelle und formale Lösungen liefert. "Das Paradigma bestimmt nicht nur, welche Probleme und Lösungsmethoden als wissenschaftlich anerkannt werden, Sogar die Art und Weise, wie ein in einer bestimmten Tradition steckender Forscher ein bestimmtes Phänomen sieht, ist durch das Paradigma bestimmt" (STEGMÜLLER, Theoriendynamik, S. 157f). Die Tragweite der Wirksamkeit solcher Paradigmen, ihr zuweilen fast weltanschaulicher Charakter und ihre für eine bestimmte Wissenschaftlergemeinschaft verpflichtende Funktion läßt sich treffend durch die Kennzeichnung von DIEDERICH (1974, S. 23) beschreiben: "Für Kuhn ist ein 'Paradigma' ja mehr als nur ein bewährter zentraler Bestandteil einer Theorie, der ohne weiteres zum Gegenstand einer Diskussion gemacht werden könnte. Er ist vielmehr eine ganze auf einem solchen Kern basierende Forschungstradition mit charakteristischer Problemsicht und beispielhaften Problemlösungen, ja, einer spezifischen Weltsicht, die in Krisenzeiten in Frage steht und meist bald von ihren Anhängern gegen ein heraufkommendes neues Paradigma verteidigt werden muß. Ein Paradigma umfaßt darum auch Kriterien für Wissenschaftlichkeit, orientiert an charakteristischen Problemstellungen und -lösungen, und es ist nicht eo ipso klar, ob Anhänger verschiedener Paradigmen sich überhaupt zu einer gemeinsamen Bewertungsbasis verstehen können".

die Rezeption von KUHN durch STEGMÜLLER die am besten systematisierte und verständlichste ihrer Art sein (Um die Wiederholung des längeren Nachweises durch die Studienausgabe der Veröffentlichung zu vermeiden, zitiere ich im folgenden mit: STEGMÜLLER, Theoriendynamik, S. ...).

Beispiele solcher Paradigmen waren oder sind die aristotelische Physik, die Elektrizitätslehre Franklins, die Kopenhagener Deutung der Quantenphysik und um ein für die Thematik der Entsubjektivierung relevantes Paradigma zu nennen, das "physikalische Weltbild" der Neuzeit, wonach wissenschaftliche Erkenntnisse unabhängig von der Person des Wissenschaftlers sein müssen. Im Bereich der Psychologie bzw. Pädagogik sind wissenschaftliche Richtungen wie Tiefenpsychologie, Gestaltpsychologie, Behaviorismus, das Konzept der kategorialen Bildung, der Curriculumansatz oder die kybernetisch orientierte Didaktik zu nennen, die nicht nur unterschiedliche gesellschaftliche Bezogenheiten und theoretische Ansätze repräsentieren, sondern auch verschiedene wissenschaftsmethodische Grundmuster der Analyse des Gegenstandsbereiches implizieren.

Zugleich beinhalten die genannten, kategorisierten Wissenschaftsrichtungen verschiedene Komplexe konkreter wissenschaftlicher Tätigkeit, konkrete normal-wissenschaftliche Tätigkeiten des Rätsellösens, deren nachträgliche Verallgemeinerung innerhalb der wissenschaftsgeschichtlichen Betrachtung erst zur Konstatierung von wirksam gewesenem oder wirksamen Paradigmen führen kann. Die historische Retrospektive zeigt allerdings, daß Paradigmen im Prozeß ihrer allmählichen Durchsetzung nicht von Anfang an mit jener Verbindlichkeit Einfluß auf die normale Wissenschaft ausübten, die ihnen vielleicht später zukam. Sie zeichnen zunächst nur einen Weg vor, dessen potentielle Erfolgsträchtigkeit sich durch ihre Bewährung für bestimmte, neue Problemlösungen legitimiert und dessen Begehung oder Ausfüllung durch die normale Wissenschaft geleistet werden soll.

Das Paradigma ist quasi ein Erfolgsversprechen, dessen Realisierung in der normalen Wissenschaft vollzogen wird. Letztlich behält es die Funktion, Erfolge in Aussicht zu stellen, während der gesamten Dauer seiner Wirksamkeit bei, wenn auch in zunehmend geringerem Maße - sei es dadurch, daß die normale Wissenschaft das Paradigma sukzessiv ausfüllt oder dadurch, daß sich bei der Anwendung des Paradigmas im Forschungsprozeß sukzessiv

Anomalien einstellen und sich im Funktionsverlust des Paradigmas die Krise andeutet, was aber gerade auf die noch vorhandene Wirksamkeit des Paradigmas als Erfolgsversprechen verweist. Die permanente, wenn auch im unterschiedlichen Ausmaß vorhandene Funktion des Paradigmas läßt sich die durch die Komplexität seiner vielfältigen Bedeutungen und Entstehungsursachen erklären, die trotz aller von ihm ausgehenden Verbindlichkeit verhindern, es als generatives System von Aussagen zu formalisieren, aus dem man eine erschöpfende Menge von wissenschaftlichen Regeln ableiten könnte.

"Wenn Wissenschaftler ein und derselben Tradition angehören, so stimmen sie in der Identifizierung des Paradigmas überein, ohne sich jedoch in der vollständigen Interpretation und Rationalisierung des Paradigmas einig sein zu müssen, ja ohne überhaupt eine solche Rationalisierung zu versuchen. Die Übereinstimmung bedeutet insbesondere, daß sie sich darin einig sind, was als legitimes Problem oder was als zulässige Lösung dieses Problems anzusehen ist" (STEGMÜLLER, Theoriendynamik, S. 158f).

Es ist dem Verdienst KUHNs zuzuschreiben, die paradigmatische Beschränkung des Konsenses innerhalb einer Wissenschaftlergemeinschaft auf die Legitimiertheit von wissenschaftlichen Problemen und die Zulässigkeit von Problemlösungen als Erfolgsaussicht für normal-wissenschaftliche Tätigkeit aufgezeichnet zu haben. Grundsätzlich dürfte der Versuch von Wissenschaftlern eines bestimmten Gegenstandsbereiches, die in der Identifizierung von Paradigmen übereinstimmen, sich über die vollständige Interpretation und Rationalisierung der Paradigmen wegen deren Normenorientiertheit, impliziten Ideologiegeprägtheit sowie deren Konstituiertheit durch vereinzelte Fälle und des subjektivistischen Einflusses der Wissenschaftler zu einigen, erfolglos bleiben. Im Bereich der Naturwissenschaften liegen aber dazu zumindest partielle und unbestreitbar brauchbare Interpretations- und Rationalisierungsansätze vor, was sich im besonderen an der auf die Naturwissenschaften bezogenen empiristischen Wissenschaftstheorie zeigt.

Im Bereich der Diagnostik in der Schule, die sich an empiristi-

schen Vorstellungen zu orientieren versucht, fehlen jedoch sogar die Bemühungen darum. So wird z.B. heute noch die wissenschaftstheoretisch unbegründbare, jedoch gesellschaftlich-historisch erklärbare Diskriminanzforderung in der Testtheorie als Teilaspekt des Entsubjektivierungsparadigmas durch lapidare Legitimationsformeln gedeckt, die sich auf die Unveränderbarkeit des selektiven Ausbildungssystems beziehen:

"Von unterschiedlichen Aufgabenschwerpunkten und entsprechend modifiziert eingesetzten Beurteilungstechniken etwa im dreigliedrigen versus Gesamtschulsystem, abgesehen kulminieren sämtliche Beurteilungs- und Beratungsansätze letztlich um die Differenzierungsproblematik. An dieser Frage und ihrer Lösung führt m.E. kein Weg vorbei, wenigstens ist bislang noch kein funktionierendes Schulsystem, das diese Problematik ohne Schaden ausklammern könnte, bekannt" (HELLER, 1974, S. 20).

Solche Legitimationswendungen müssen offensichtlich für den Umstand einspringen, daß man sich über die Interpretation und Rationalisierung der Diskriminanzforderung für den diagnostischen Forschungsprozeß nicht nur nicht einigen kann, sondern beides nicht versucht wird. Letzteres belegt FRICKE (1972) durch die zirkuläre Einführung des Diskriminanzprinzips, ohne daß er sich dessen bewußt ist. Während er auf Seite 25 zugibt, daß es sich bei der individuell unterschiedlichen Schulleistung um Eigenschaften von Schülern, um *e r f u n d e n e A b s t r a k t i o n e n* handelt, die der näheren Definition durch den Beobachter bedürfen - und zwar hier der Definition im Sinne eines metrischen Begriffes -, fordert er auf Seite 35 die Grundannahme, "daß die Testwerte in einer systematischen Weise mit den Fähigkeiten variieren, daß z.B. einem 'mehr' bei den Testwerten ein 'mehr' bei den Konstrukten entspricht ...". Was definiert werden soll, muß als Grundannahme vorausgesetzt werden, damit es definiert werden kann, nämlich die metrische Struktur eines Konstruktes oder einer Abstraktion.

Unter solchen Voraussetzungen trifft KUHNs Kennzeichnung der Wissenschaftsentwicklung in jenem mißverstandenen Sinne zu, der im "Irrationalismus"- und "Subjektivismus"-Vorwurf von POPPER enthalten ist, indem dem normalen Wissenschaftler bornierter

Dogmatismus unterstellt wird. POPPER glaubt aus der KUHNschen Unterscheidung zwischen dem normalen Wissenschaftler und dem in der noch zu beschreibenden, außerordentlichen bzw. revolutionären Wissenschaft tätigen Forscher die Konsequenz ziehen zu müssen, daß ersterer sich durch eine unkritische Einstellung und Überangepaßtheit an die herrschende Lehrmeinung auszeichne. "Nach meiner Auffassung ist der 'normale' Wissenschaftler, wie KUHN ihn beschreibt, eine Person, die einem leid tun sollte" (POPPER, 1970, S. 52).

Abgesehen von einigen mißverständlichen und z.T. auch widersprüchlichen Ausführungen von KUHN, die die POPPERsche Fehlinterpretation der genannten Unterscheidung naheliegen und deren kritische Diskussion KUHN zu Modifikationen und Korrekturen zwang, erscheint eine andere Deutung angemessener, die aber für die empiristisch orientierte Diagnostik in der Schule zu den gleichen negativen Konsequenzen führt. In seinem Rekonstruktionsversuch der KUHNschen Position entkräftet STEGMÜLLER den Dogmatismusvorwurf gegenüber dem normalen Wissenschaftler dadurch, daß dessen monierte Angepaßtheit sich auf einen engen Teilbereich theoretischer Vorstellungen bezieht: unter der Voraussetzung, daß sich das Gemeinsame in einer Gruppe von normalen Wissenschaftlern durch die Verfügung über ein und dieselbe Theorie (der Aspekt, demgegenüber sie "unkritisch eingestellt" sind) angeben läßt, obwohl die theoretischen Überzeugungen, Vermutungen und Hypothesen ständig wechseln, und "daß der 'theoretische Aspekt' des Verfügens über eine Theorie im S t r u k t u r - k e r n , also in einer mathematischen Struktur und nicht in einem System von Aussagen, verankert werden wird" (STEGMÜLLER, Theoriendynamik, S. 181), verbleibt dem normalen Wissenschaftler noch genügend Spielraum - ohne daß er einem leig tun kann -, kreative Impulse zu liefern und kritische Wissenschaft zu betreiben.

Man mag allerdings bezweifeln, ob der STEGMÜLLERsche Rekonstruktionsversuch Licht in das Dunkel der Integrations- und Rationalisierungsmöglichkeit der in der empiristisch orientierten Dia-

gnostik in der Schule geltenden methodischen Paradigmen bringt. Denn wo findet man hier mathematische Strukturen als Strukturkerne von Theorien, die den diagnostischen Gegenstandsbereich mit der Aussicht auf Gewinnung von zuverlässigen und gültigen Urteilen über Schüler hinreichend systematisieren? Gibt es in der Diagnostik in der Schule ein analysiertes Gegenstandssystem, von dem entsprechend der SNEEDschen Interpretation der Funktion von Strukturkernen zu behaupten ist, es besitze eine bestimmte mathematische Struktur?

Die bestehenden Strukturierungsbemühungen für Urteilsdaten über Schüler durch die klassische Testtheorie beinhalten eine Reihe von syntaktischen und semantischen Schwierigkeiten, die sich als Folge eines völlig ungeklärten aber dennoch verwendeten Meßmodells ergeben, dessen Interpretation und Rationalisierung für den Forschungsprozeß als Paradigma in der Bedeutung eines Strukturkerns nicht möglich ist (vgl. TACK, 1970). Ein typisches Beispiel bildet das sog. "Attenuation Paradox". Es besteht darin, "daß die Erhöhung der Reliabilität eines Tests unter bestimmten Bedingungen eines zwangsläufige Verringerung der Validität zur Folge hat ... Die Ergebnisse legen den Schluß nahe, daß das unerwünschte Verhalten der Validität aus der Natur des Korrelationsbegriffes folgt und daher mit den grundlegenden Eigenschaften der klassischen Testtheorie zusammenhängt" (FISCHER, 1974, S. 145). Aber auch die neueren Formalisierungsbestrebungen lassen vorerst die Schwierigkeiten der Meßproblematik beiseite, wenn sie die von RASCH (1966) thematisierte Bedingung für die Wissenschaftlichkeit von Aussagen nicht einzulösen beginnen, wonach Wissenschaftlichkeit nicht darin mündet, "überprüfbare" Tatsachenfeststellungen zu machen, sondern Vergleiche zwischen verschiedenen Objekten durchzuführen, die ihrerseits "objektiv" sein sollen. "Eine wissenschaftstheoretische Analyse hat daher in erster Linie den Begriff des V e r g l e i c h e s zu spezifizieren und dem Begriff O b j e k t i v i t ä t eine präzise Bedeutung zu geben. In der Folge ist dann zu untersuchen, unter welchen Bedingungen objektive Vergleiche möglich sind" (FISCHER, 1974, S. 408).

Es deutet sich in diesen Bemerkungen zumindest die Tendenz an, die wissenschaftstheoretische Vorleistung der Objektbestimmung vor dem formalisierten Vergleich der Objekte als notwendig zu erkennen. Ob das auf traditionell empiristischem Wege durch die Vernachlässigung des Entstehungszusammenhanges von Gegenständen und Gegenstandsbereichen durchzuführen ist, wird selbst innerhalb der neueren Entwicklungen der Wissenschaftstheorie bestritten. Auf jeden Fall wäre damit ein erster, formaler Schritt in Richtung auf eine empirisch orientierte Verwissenschaftlichung der Diagnostik in der Schule erfolgt. Ohne die Konkretion dieser Umorientierung bleibt die paradigmatische Beschränkung des Konsensus auf die Verwendung eines noch so differenzierten Methodenarsenals ein beliebig interpretierbares Vorgehen. Dann läßt sich auch das empiristisch orientierte Vorgehen in der Diagnostik in der Schule mit jedem Paradigmabegriff kennzeichnen, den POPPER dem KUHNschen Ansatz unterstellt und durch den der normale Wissenschaftler als bornierter Dogmatiker ausgewiesen wird.

Immerhin wäre ein derartig "irrationaler" Zustand des Wissenschaftsbetriebes - und damit kehre ich zur Darstellung der KUHNschen Position zurück - die Vorbedingung für einen zu erwartenden Paradigmenwechsel. Denn das Bedürfnis nach einer zunehmenden Formalisierung und Reglementierung wissenschaftlicher Tätigkeit entsteht immer erst dann, wenn der Verdacht der Unzuverlässigkeit des Paradigmas aufkommt. Dieser Verdacht begründet sich auf das gehäufte Auftreten von Anomalien, die durch die herkömmlichen Methoden nicht mehr zu bewältigen sind. "Erst die Perioden unmittelbar vor der Einführung neuer Paradigmata sind regelmäßig von Debatten über die gültigen Methoden, Probleme und Lösungsmuster begleitet, also diejenigen Perioden, in denen Paradigma zunächst angegriffen und dann einem Wechsel unterzogen werden" (STEGMÜLLER, Theoriendynamik, S. 159). Zur Veranschaulichung der aktuellen Relevanz des Zitats für die empiristisch orientierten Sozialwissenschaften und der Sozialwissenschaften im allgemeinen kann man auf eine Bemerkung von KUHN selbst zu-

rückgreifen:

"... das Jahr einer Gemeinschaft, die sich überwiegend aus Sozialwissenschaftlern zusammensetzte, konfrontierte mich mit unerwarteten Fragen über die Unterschiede zwischen solchen Gemeinschaften und jenen der Naturwissenschaftler, in denen ich ausgebildet worden war. Insbesondere war ich überrascht von der Zahl und dem Ausmaß der offenen Meinungsverschiedenheiten unter den Sozialwissenschaftlern über das Wesen der g ü l t i g e n P r o b l e m e u n d M e t h o d e n" (Sperrung vom Verfasser) (KUHN, 1973, S. 10).

Allgemein steht die normale Wissenschaft in der kriseninduzierenden Phase vor dem Rätsel einer wachsenden Menge von ungelösten Rätseln. Eine systematische Charakterisierung derjenigen Anomalien, die ein tradiertes Paradigma herausfordern können, fehlt bei KUHN. STEGMÜLLER faßt die Funktion der Anomalien in der Wissenschaftsentwicklung dahingehend zusammen, daß sie je nach der Grundsätzlichkeit der Annahmen, die sie in Frage stellen, sowie je nach dem Stärkegrad der Abweichung von dem, was der traditionellen Theorie gemäß zu erwarten war, in den Forschern eine mehr oder weniger tiefe Beunruhigung hervorrufen. Mit dem Fortdauern dieses Prozesses kommt es zur Krise und - was für die gegenwärtige Verwissenschaftlichung der Diagnostik in der Schule unter Berücksichtigung der Tatsache besonders typisch ist, daß man hier nicht über d i e Theorie, sondern nur über theorieähnliche Fragmente verfügt - divergierende Ausarbeitungen der Theorie beginnen zu wuchern und die Regeln der normalen Wissenschaft werden in zunehmenden Maße aufgeweicht. Solche divergierenden Ausarbeitungen heißen dann später ad-hoc-Anpassungen (ebd., S. 163).

Damit ist der Zeitpunkt für das in Erscheinungtreten der außergewöhnlichen bzw. revolutionären Wissenschaft gekommen. Sie konkretisiert sich in der Ankündigung und allmählichen Durchsetzung von radikal veränderten Paradigmen, die, an spezifischen Beispielen entwickelt, einen breiten Bereich der existierenden Anomalien zu subsumieren in der Lage sind und die *conditio-sine-qua-non* für die Absetzung der mit ihnen inkommensurablen, tradierten Paradigmen bilden.

Die Annahme von der Verdrängung eines Paradigmas durch ein anderes, sowie die These von der Inkommensurabilität der rivalisierenden Paradigmen hat seit KUHNs Veröffentlichung einen heftigen Disput ausgelöst, zumal KUHN selbst keine systematischen Bedingungen für die Notwendigkeit und Möglichkeit des Beginns revolutionärer Wissenschaft ausgeführt hat und ebensowenig für das Wirksamwerden sozialpsychologischer Faktoren, die als Motivations- und Kognitionsstrukturen einzelne Wissenschaftler dazu bewegen, sich mit dem Anomalienproblem zu beschäftigen. Die Auseinandersetzung mit der Inkommensurabilitätsannahme mußte deswegen erfolgen, weil, wenn es richtig wäre, daß zwischen verdrängender und verdrängter Theorie kein Vergleich möglich ist, dann keine rationalen Argumente zwischen den beiden Theorien entscheiden können. Dann wird in der Tat KUHNs Auffassung bestätigt, daß nicht die besseren Begründungen, die besseren Argumente und die bessere Bewährung an der Erfahrung für eine Theorie sprechen, sondern beide Theorien sich in Konkurrenz zueinander befinden und im Kampf gegeneinander jene obsiegt, die mehr "Rätsel" umfassender zu lösen vermag. Dies würde auch mit der Behauptung konform gehen, daß ein tradiertes Paradigma mit dem Aussterben der es vertretenden Wissenschaftler seine Funktion verliert und daß sich nach einem Paradigmawechsel die Welt selbst verändert hat.

In dieser Behauptung äußert sich KUHNs subjektivistische und idealistische Auffassung, die ihn nicht nur daran hindert, das Aufkommen von Anomalien zu begründen, sondern die entsprechenden Bedingungen überhaupt aufzusuchen. Er wird dadurch inkonsequent und gerät in Widersprüche. Denn einerseits leugnet er die Möglichkeit der rationalen Rekonstruktion der Wissenschaftsentwicklung sowie ihre Interpretation als kummulative Anhäufung erreichten Wissens. Er stellt ihr den Kampf der Theorien und Paradigmen gegenüber, die zu einer wissenschaftlichen Revolution führen. Andererseits kann er aber nicht über die objektive Vorhandenheit provozierender Anomalien hinweggehen und scheint auch objektivierbare Kriterien des Theorienvergleiches zu ken-

nen, die der Inkommensurabilitätsthese zumindest in ihrer Allgemeinheit widersprechen, z.B. Voraussagekraft, Belastung durch Anomalien, Bewahrung schon erreichter Problemlösungen und Lösung zuvor vergeblich angegangener Probleme.

Nun gibt es m.E. zwei Wege, die Inkommensurabilitätsthese zu klären und in einem eingeschränkten Sinne beizubehalten:

Entweder, man schließt sich der umfassenderen Theorie- und Paradigmaauffassung von KUHN an und bezieht traditionell außerwissenschaftliche Konstitutionskriterien in die Betrachtung der Wissenschaftsentwicklung mitein. Das würde aber bedeuten, den Bereich der Konstitutionsmomente ziemlich weit zu fassen, über die sozialpsychologischen und soziologischen Faktoren hinauszugehen und die Abhängigkeit der Wissenschaftsentwicklung von der gesellschaftlichen Entwicklung zu hinterfragen (vgl. BÖHME, v.d. DAELE & KROHN, 1974). KUHN tastet sich meines Wissens an einer einzigen Stelle in seinem Buch an dieses Problem heran, wo er die Bedingungen für das Aufkommen von Anomalien exemplifiziert, ohne jedoch die gesellschaftliche Relevanz explizit zu thematisieren.

"Oder aber eine Anomalie ohne sichtbare grundlegende Bedeutung kann, wie in der Kopernikanischen Revolution, eine Krise hervorrufen, wenn die Anwendungen, die sie verhindert, für die Praxis besonders wichtig sind, in diesem Fall für die Aufstellung eines Kalenders oder für die Astrologie. Oder die Entwicklung der Normalen Wissenschaft kann - wie in der Chemie des achtzehnten Jahrhunderts - eine Anomalie, die vorher nur eine Belästigung gewesen war, in die Ursache einer Krise verwandeln: Das Problem der Gleichgewichtsrelation erhielt durch den Fortschritt bei den pneumatisch-chemischen Verfahren einen ganz anderen Status" (KUHN, 1973, S. 116f).

Es zeigt sich also, daß sich die Beständigkeit von Anomalien bzw. ihr Wichtigwerden im Forschungsprozeß durchaus objektiv begründen läßt. Desgleichen könnten die Ursachen für den Beginn der Auseinandersetzung von vereinzelt Wissenschaftlern mit wissenschaftlichen Rätseln, sowie die Entwicklung eines obsiegenden Paradigmas auf der Ebene der gesellschaftlichen Relevanz soweit untersucht werden, daß die Entscheidung im Kampf zwischen rivalisierenden Theorien oder Paradigmen nicht länger das Resultat

tat eines mystischen Kräftespiels bleibt und nicht nur an den Tod von Wissenschaftlern gebunden ist. Nur durch die Außerachtlassung des gesellschaftlichen Entstehungszusammenhanges von Paradigmen im KUHNschen Sinne und mit dem Anspruch eines formalen Rationalismus kann die Wissenschaftsentwicklung als "irrationaler" Prozeß erscheinen.

Oder aber - und das ist die zweite Interpretationsmöglichkeit der Inkommensurabilitätsthese -, man folgt dem Rekonstruktionsversuch von STEGMÜLLER und SNEED und begreift Paradigma als Strukturkern einer Theorie mit einer durch die Angabe typischer Fälle umrissenen Menge von intendierten Anwendungen. Dann entspricht der behaupteten Inkommensurabilität von Paradigmen die Verschiedenheit der Strukturkerne der beiden betreffenden Theorien. Dann ist aber die Unmöglichkeit des Theorienvergleichs nur auf einen formalisierten Aspekt beschränkt, der Leistungsvergleich der gesamten Theorien selbst nur wenig berührt. Abgesehen davon, daß diese Art der wissenschaftstheoretischen Rekonstruktion des Paradigmbegriffes differenziert durchstrukturierte Theorien voraussetzt und hohe Ansprüche an eine empiristische Theorienkonstruktion und -entwicklung stellt, werden beim Leistungsvergleich von Theorien, die auch das Ergebnis der normalen wissenschaftlichen Tätigkeit einschließen, wieder alle Probleme aktuell, die in den Bereich der Bestätigungs- und Testfragen gehören.

Hier erweist sich allerdings wiederum eine KUHNsche Kritik als bedeutsam, die die logische Begründung seiner Inkommensurabilitätsthese bildet und die im wissenschaftstheoretischen Rekonstruktionsversuch zu Schwierigkeiten führt. Bezogen auf den noch globalen Begriff des Paradigmas will KUHN seine These durch die Nichtexistenz einer neutralen Beobachtungssprache belegen, die für Tests und Prüfungen von miteinander konkurrierenden Paradigmen notwendig wäre. Der Gedanke an die Existenz einer theorie-neutralen Beobachtungssprache ist für ihn illusionär.

Die Begründung ist jedoch irreführend und inkonsequent, weil er in seinem nach wissenschaftstheoretischen Kriterien undifferen-

zierten Paradigmbegriff keine explizite Unterscheidung zwischen Theorienkonstruktion und -überprüfung bzw. theoretischer Sprache und Beobachtungssprache trifft und auch nicht kann, da für ihn das Obsiegen eines Paradigmas nichts mit empirischer Bewährung zu tun hat. Der Hinweis auf die Unmöglichkeit einer theorieneutralen Beobachtungssprache wird aber dann relevant, wenn der Leistungsvergleich von Theorien - wie im wissenschaftstheoretischen Rekonstruktionsversuch - auch auf den intendierten Anwendungsbereich von Theorien bezogen wird, der gewöhnlich durch paradigmatische Beispiele festgesetzt ist. Da für die Formulierung von paradigmatischen Beispielen beobachtungssprachliche Terme herangezogen werden, stellt sich natürlich die Frage nach einer neutralen Sprache zum Vergleich der intendierten Anwendungsbereiche solange, wie die Konstituiertheit der Beobachtungssprache unberücksichtigt bleibt. Dadurch wäre aber KUHNs Inkommensurabilitätsthese auf dem wissenschaftstheoretischen Verarbeitungsniveau des Paradigmbegriffes partiell bestätigt.

Mit den beiden Explikationsansätzen der Inkommensurabilitätsthese möchte ich die zusammengefaßte Vorstellung der Position KUHNs beenden. Sie wurde an das Ende der Skizzierung des Paradigmbegriffes gestellt, weil sich in ihr der "Irrationalismus"-Vorwurf gegenüber der Wissenschaftsentwicklung konzentriert und sich daran die Betrachtung der kritischen Einwände von seiten der Wissenschaftstheoretiker in inhaltlicher Kontinuität anschließen läßt. In Zusammenhang damit soll die Frage angesprochen werden, ob sich aus der wissenschaftstheoretischen Erörterung und Kritik der KUHNschen Auffassung Richtlinien für die Standortbestimmung der empiristisch orientierten Diagnostik in der Schule ergeben.

1.4.3 Kritische Einwände gegenüber KUHN und ihre Bedeutung für die Entsubjektivierung der Diagnostik

Es wurde vorher bereits darauf hingewiesen, daß bestimmte kritische Einwände gegenüber KUHN legitim zurückweisbar sind, wenn seine Auffassung über die Wissenschaftsentwicklung auf die Naturwissenschaften - im besonderen auf die exakten Naturwissenschaften - bezogen werden. Sie können hingegen volle Gültigkeit haben, wenn der Forschungsprozeß im Bereich der Diagnostik unter die Lupe genommen wird. Dabei muß allerdings die erste Explikationsweise der Inkommensurabilitätsthese völlig unberücksichtigt bleiben, da weder in den Naturwissenschaften noch in den empiristisch orientierten Sozialwissenschaften die Reflexion des Zusammenhanges von wissenschaftlichem Fortschritt und gesellschaftlicher Entwicklung stattfindet. Zudem liegt die wissenschaftstheoretische Rezeption der KUHNschen Konzeption auf einer Ebene mit dem Anspruch der empiristisch orientierten Sozialwissenschaften, sodaß nur von dort her auszumachen ist, ob letztere dem "Irrationalismus"- bzw. "Subjektivismus"-Verdikt in der Interpretation POPPERS verfallen oder die Aussicht besteht, deren Theoriefragmente einem empiristisch ausgerichteten Leistungsvergleich zu unterziehen.

Einer der wichtigsten, insbesondere für die Betrachtung der verwissenschaftlichten Diagnostik in der Schule bedeutsamen, kritischen Einwände stammt von SCHEFFLER (1967). Er wendet sich gegen KUHNs Behauptung, daß zwischen Vertretern verschiedener Paradigmen nicht einmal eine gemeinsame Wissenschaftssprache existiert, ja, daß sogar innerhalb einer auf ein bestimmtes Paradigma eingeschworenen Wissenschaftlertgemeinschaft kein Konsensus bezüglich der Hypothesen, Vorstellungen, Aussagen und wissenschaftlichen Vorgehensregeln besteht.

Die Behauptung steht in krassem Gegensatz zu der Auffassung, den Forschungsprozeß entsubjektivieren zu können. Unter Berufung auf die REICHENBACHsche Unterscheidung von Entdeckungs-

und Rechtfertigungszusammenhang von Theorien setzt SCHEFFLER die Unabhängigkeit der Kriterien für den objektivierten Vergleich von Theorien entgegen. Zwar könnten die Anhänger verschiedener Paradigmen einander nicht von ihrem eigenen Forschungsansatz überzeugen, sie hätten aber allgemein anerkannte Beurteilungskriterien pragmatischer Art zur Hand, weshalb Inkommensurabilität der Paradigmen nicht zugleich deren Unvergleichbarkeit bedeute. Bei dieser Argumentation stützt sich SCHEFFLER auch auf eine scheinbare Inkonsistenz von KUHN, indem er dessen Zugeständnis benutzt, daß Beurteilungskriterien in Paradigmenauseinandersetzungen durchaus existieren - die bereits erwähnten Kriterien der Voraussagekraft, Belastung durch Anomalien usw.

Das Bekenntnis KUHNs zu diesen Kriterien ist unbestritten; es kommt aber darauf nicht an. Das Problem besteht vielmehr darin, daß KUHN die übereinstimmende Interpretation und Rationalisierung der Kriterien durch Wissenschaftler verschiedener Provenienz für den Forschungsprozeß in Zweifel zieht und nicht ihre Existenz. Dies wird auch durch die Neufassung seines Paradigmakonzeptes unter dem Einfluß der begonnenen Auseinandersetzung mit der ursprünglichen Fassung bestätigt. KUHN bezeichnet dort die erwähnten Kriterien zur Einschätzung von Paradigmen, neben anderen wie Fruchtbarkeit und Einfachheit, als übergreifende Werte und geht nicht davon aus, daß "die gemeinsame Anerkennung dieser Werte eine Wissenschaftler-Gemeinschaft auch im konkreten Fall zu einem einhelligen Urteil führt. Vielmehr machen es die Unsicherheit bei der Subsumption konkreter Fälle unter die Werte wie auch gerade die Mehrzahl der Werte dem einzelnen Forscher möglich, bei gleicher Wertbasis durch Urteilsdifferenzen und verschiedene Rangfolgen der Werte verschiedene Präferenzen zu setzen" (DIEDERICH, 1974, S. 24).

Von da aus gesehen, muß SCHEFFLERS Einwand auf den Aspekt der übereinstimmenden Interpretation und Rationalisierung von Paradigmen für den Forschungsprozeß bezogen werden. Dazu macht er aber keine näheren Ausführungen. Er beschränkt sich letzten-

endes nur auf die Angabe der Problemlösungsrichtung, wonach die Frage der Kommunikabilität der Kriterien als Gegenstand einer gesonderten Analyse auf metatheoretischer Ebene behandelt werden muß. Man kann unschwer vorhersagen, daß der vorgeschlagene Lösungsweg zu einem konventionalistischem Regreß führt. Überdies muß SCHEFFLER aus der Unterscheidung von Entdeckungs- und Rechtfertigungszusammenhang die Konsequenz ziehen, ersteren durch die pragmatische Verallgemeinerbarkeit von dann entsubjektivierten Kriterien des Leistungsvergleiches von Theorien in den Forschungsprozeß wieder hereinnehmen und den Subjektivismus durch die Hintertür wieder hereinzulassen.

Beim Rekonstruktionsversuch für den Paradigmbegriff durch STEGMÜLLER tritt das gleiche Problem dort auf, wo es um die Festlegung der Menge der intendierten Anwendungen einer Theorie geht, die als Modell des Strukturkerns und Strukturrahmens gelten sollen. Für die Frage der Kommunikationsmöglichkeit zwischen den Wissenschaftlern ist die auf drei Ebenen durchgeführte Rationalisierung des KUHNschen Paradigmbegriffes wichtig. Als Bedeutungen kommen in Frage:

1. "Mögliche Objekte präziser logisch-wissenschaftstheoretischer Rekonstruktion" (STEGMÜLLER, Theoriendynamik, S. 203). Dabei handelt es sich um jenen Teil der mathematischen Struktur einer Theorie, die im Rahmen der normalwissenschaftlichen Tätigkeit unverändert bleibt.
2. "Die durch paradigmatische Beispiele bestimmte Menge I der intendierten Anwendungen" (ebd., S. 204). Die Einzelheiten der wissenschaftstheoretischen Einführung von Beispielsmen-gen müssen hier nicht referiert werden. Zur Veranschaulichung sei auf den von STEGMÜLLER zitierten Fall der klassischen Partikelmechanik verwiesen, "für welche noch immer diejenigen paradigmatischen Beispiele gelten, die bereits zu NEWTONs Zeiten als Anwendungen seiner Theorie erkannt worden sind, nämlich: das Planetensystem und verschiedene Teilsysteme davon (Erde - Mond, Jupiter - Jupitermonde); einige Kometen; die Körper im freien Fall nahe der Oberfläche der Erde; die Gezeiten; die Pendel" (ebd., S. 200).
3. "Diejenigen Komponenten des Paradigmas im Kuhnschen Sinn, die wissenschaftstheoretisch nicht faßbar sind, da sie in die Psychologie und Soziologie der Forschung hineingehören" (ebd., S. 204).

Die Ausklammerung der dritten, der psychologischen und soziolo-

gischen Erforschung vorbehaltenen Restkategorie als Teilaspekt der Wirksamkeit von Paradigmen ist natürlich aus wissenschaftstheoretischer Perspektive zu erwarten und verleiht den impliziten Subjektivismus der Gesamtposition. Sie wird denn auch von STEGMÜLLER als störender Faktor in der Kommunikation zwischen Wissenschaftlern anerkannt und in Entsprechung zum Paradigmenbegriff von KUHN gesehen. Ebenso konkretisiert sich damit noch einmal die Unterscheidung von Entdeckungs- und Rechtfertigungszusammenhang, die STEGMÜLLER zwar für die Theoriebildung als gegenstandslos betrachtet, nicht aber für die Hypothesenbildung. Entscheidender ist innerhalb der immanenten Betrachtung die Behauptung, daß mit der Festlegung der intendierten Anwendung einer Theorie durch paradigmatische Beispiele eine Rationalisierung des Paradigmas durchzuführen sei, die eine hinreichende und willkürfreie Kommunikation zwischen Wissenschaftlern schaffe. Diese Behauptung ist erkenntnistheoretisch außerordentlich problematisch, da die Festlegung der Beispielsmenge einer erkenntnisfunktionalen Begründung bedarf, die jedoch paradigma- und theorieimmanent nicht mehr einzuholen ist. Sie kann daher auch nur in der historischen Retrospektive eine Bestätigung finden und zwar nur dort, wo sich ein nachträglich rekonstruierbarer Anwendungsbereich entwickelt hat ⁶⁾.

STEGMÜLLER charakterisiert das Dilemma mit der in Kauf zu nehmenden Vagheit im Entwicklungsprozeß von Theorien, die aber keinesfalls ein vernunftwidriges Verhalten der Wissenschaftler widerspiegeln und im Gegensatz dazu zur Präzisierung der Forschungsbemühungen herausfordern. Solange er dabei an den Gegenstandsbereich der theoretischen Physik denkt, mag die Vagheit

6) Nicht umsonst bezieht sich STEGMÜLLER in seiner Explikation des KUHNschen Paradigmenbegriffes auf physikalische Theorien, die in hohem Maße differenziert sind. Ob die Kommunikation von Sozialwissenschaftlern auf dieser Ebene ablaufen kann, läßt sich nicht nur durch die bestehende Sprachverwirrung und durch das Fehlen vorauszusetzender, formalisierter Theoriestrukturkerne widerlegen, sondern vor allem durch die Unfähigkeit, paradigmatische Beispiele anzuführen, die sich aus einer erst zu leistenden Gegenstandsanalyse ergeben könnten.

nicht als Hinderungsgrund und Irrationalitätsfaktor für die Kommunikation zwischen den Wissenschaftlern erscheinen, wenn-
gleich ihre Vorhandenheit selbst noch einmal aus der Besonder-
heit der Festlegung der intendierten Anwendungsbereiche begrif-
fen werden müßte. Die darin implizierten, gesellschaftlichen
Bedingungen als Verwendungskontext wissenschaftlicher Tätigkeit
wären in ihrer historischen Entwicklung ebenso wie die darin an-
gelegten, aber heute erst rekonstruierbaren, theoretischen Struk-
turkerne zu untersuchen. Aufgabe einer solchen Untersuchung wä-
re es, zum einen die auch für die Naturwissenschaften existie-
renden, vom historischen Stand der gesellschaftlichen Entwick-
lung bestimmten, wissenschaftlichen Erkenntnisbeschränkungen
herauszuarbeiten und zum anderen die Übertragungsprobleme zu
reflektieren, die aus der Forderung resultieren, Wissenschaft
schlechthin zu formalisieren und zu entsubjektivieren.

Bevor dies nicht geleistet ist, scheitert die Kommunikation
zwischen Vertretern verschiedener Paradigmen an den von KUHN
benannten "Irrationalitäts"-Barrieren, ohne daß sich die Betrof-
fenen dessen bewußt sind. Dann modellieren u.U. verschiedene
Wissenschaftlergemeinschaften gleiche oder unterschiedliche For-
malstrukturen in abweichender Weise und betrachten ihren Disput
als Theorienstreit. Dann trifft auch zu, was STEGMÜLLER unter
der KUHNschen These der Kommunikationsunmöglichkeit versteht:
"Man kann nicht mit Hilfe der mathematischen Struktur einer
Theorie diese Struktur mit einer davon verschiedenen vergleichen.
Doch wird solches in Krisenzeiten von den beteiligten Wissen-
schaftlern immer wieder versucht" (ebd., S. 174).

Dies ist nun der Anknüpfungspunkt, aus dem Verhältnis zwischen
der relativ globalen Auffassung KUHNs gegenüber den Kommunika-
tionsschwierigkeiten unter Wissenschaftlern und der wissenschafts
theoretisch interpretierten Auffassung zu einer Kennzeichnung
der Verwissenschaftlichungstendenzen in der Diagnostik in der
Schule zu gelangen. Es wird zu zeigen sein, daß die einzelnen
Ansätze zur methodischen Objektivierung der Schülerbeurteilung
auf formalen Strukturen aufbauen, für die keine, durch paradigm-

matische Beispiele festgelegte, intendierte Anwendungsbereiche zu finden sind.

So haben z.B. die beiden gegensätzlichen Formalisierungsbestrebungen der normorientierten bzw. kriterienorientierten Testung keine paradigmatischen Modellinterpretationen erfahren, über die sich die Diagnostiker verständigen könnten. In der klassischen Testtheorie, die der Konstruktion normorientierter Tests zugrundeliegt, setzt die Anwendung der Formel für den Standardmeßfehler voraus, daß die Verteilung der Fehler um den wahren Wert eines Individuums normal ist und daß die Größe des Standardfehlers von der Größe des wahren Wertes, für den die Schätzung beabsichtigt ist, unabhängig ist. "Beide Annahmen sind Gegenstand häufiger Diskussion gewesen. Die letztere, daß die Fehlerwerte vom wahren Wert unabhängig sind, ist besonders kritisiert worden" (MAGNUSSON, 1969, S. 89). Bei kriterienorientierten Tests, die den Abstand des aktuellen Lernerfolges zu festgelegten Lernzielen zu erfassen suchen, sind die klassischen Methoden zur Gütebestimmung der Tests wegen der reduzierten Varianz der Beurteilungsergebnisse nicht anwendbar. Neuere Ansätze zur Konstruktion und Gütebestimmung von kriterienorientierten Tests sind in ihrer Entwicklung noch nicht über ein Anfangsstadium hinausgekommen. "Alle die genannten Verfahren zur Standardisierung des Itemkonstruktionsprozesses sind problematisch. Sie sind in ihrer Anwendung meist sehr umständlich und sie versagen im allgemeinen bei komplexen Sachverhalten" (BÜSCHER, 1974, S. 154).

Außerdem repräsentieren beide Konzeptionen keine Strukturkerne von Theorien im Sinne einer mathematischen Struktur, die als Bestandteil der Theorie unverändert bleiben kann. Im Gegensatz dazu wird ja gerade versucht, für den Formalismus einen theoretischen Rahmen zu finden, in dem ersterer sich vielleicht als stabil erweisen könnte. Das würde insgesamt bedeuten, daß die wissenschaftstheoretische Rationalisierung des Entsubjektivierungsparadigmas - bezogen auf die beiden Konzeptionen - für die Diagnostik in der Schule nicht gültig ist.

Andererseits wird aber damit im Verwissenschaftlichungsprozeß

der Diagnostik so verfahren, als seien beide Konzeptionen mit von Fall zu Fall unterschiedlicher Richtung erfolgsversprechende Objektivierungsrichtlinien für die normalwissenschaftliche Tätigkeit und als könnten sie gegeneinander ausgespielt werden. Berücksichtigt man außerdem - und dies gilt bevorzugt für die Methodik der normorientierten Testung -, daß durch sie eine Reglementierung des Forschungsprozesses vorgenommen wird, deren Nichteinhaltung weitgehend unkritisch mit dem Urteil "methodisch unsauber" belegt wird, dann scheint hier ein Paradigma wirksam zu sein, daß eher der KUHNschen Interpretation entspricht und zu "irrationalen" Verwissenschaftlichungsprozessen führt. Es scheint daher nichts anderes übrig zu bleiben, als den Ursprung solcher methodischer Paradigmen sowie ihre wirksame Verbindlichkeit in außerwissenschaftlichen Bereichen zu lokalisieren, nämlich in der gesellschaftlich durchgesetzten Selektion und in seiner Form des organisierten Lernens, bzw. in den außerwissenschaftlichen Bereichen, in denen die Konzeption der differentiellen Psychologie und der ihr entsprechende methodische Komplex ihre Grundlagen haben.

Zuvor ist aber noch ein Problem zu behandeln, das für den Aspekt der methodischen Objektivierung der Schülerbeurteilung von zentraler Bedeutung ist. Es geht um die Frage nach der Funktion methodologischer und methodischer Regeln im Forschungsprozeß und der daraus resultierenden Konsequenzen für die Interpretation dadurch erhaltener Untersuchungsergebnisse in den empiristischen Wissenschaften. Solche Regeln formalisieren einen Großteil normalwissenschaftlicher Tätigkeit wie das Aufstellen, Verbessern und Preisgeben von Hypothesen, die Einführung von wissenschaftlichen Begriffen auf unterschiedlichen Bedeutungsniveaus die Untersuchung von Nebenbedingungen für die Gültigkeit von Gesetzen usw. KUHN hat diesen Problemkreis aus seinen Betrachtungen faktisch ausgeklammert und unter dem globalen Namen "Rätsellösen" zusammengefaßt. Sie schienen ihm als Gegenstand der Untersuchung vernachlässigbar zu sein, weil er ihre Rolle im Prozeß der wissenschaftlichen Revolution als unbedeutend ein-

schätzte und für ihn ihre Verwendung in keinem rationalen Bezug zum historisch wirksamen Paradigma stand.

Die Auffassung der letztlich willkürlichen Anwendbarkeit von Regeln der wissenschaftlichen Tätigkeit im Bereich der Normalwissenschaft ist der Hintergrund für die Rezeption von KUHN durch POPPER und auch LAKATOS. Danach ist der normale Wissenschaftler ein kritikloser Dogmatiker, der seine Vorstellungen kaum einer Prüfung unterzieht, wohingegen der revolutionäre Wissenschaftler einem religiösen Fanatiker ähnelt, der sein Bekehrungserlebnis mit allen Mitteln der Überredung und Propaganda durchzusetzen versucht. Diese Rezeption ist nicht so leicht von der Hand zu weisen, wenn man an die bereits zitierte Legitimation des konstruktivistischen Standpunktes durch WESTMEYER denkt.

Der Denkanstoß durch die Herausforderung durch KUHN muß dabei so stark gewesen sein, daß LAKATOS (1970), um sich der "Irrationalismus"-Interpretation der Wissenschaftsentwicklung entgegenzustellen, den Gedanken einer normativen Methodologie entwickelt hat, die auf eine Reglementierung der wissenschaftlichen Tätigkeit hinausläuft. LAKATOS, dem es vor allem darum ging, die Relevanz wissenschaftstheoretischer Rekonstruktion für die Wissenschaftsentwicklung aufzuzeigen, will mit seiner normativen Methodologie dem Wissenschaftler quasi vorschreiben, welche Forschungswege er zu begehen hat und welche von ihm zu meiden sind.

Seine Reaktion auf KUHN ist für die vorliegende Betrachtung der verwissenschaftlichten Diagnostik in der Schule als Teilbereich empiristisch orientierter Sozialwissenschaft wichtig, weil sie im Hinblick auf die Intention zu verstehen ist, die Wissenschaftsgeschichte in Teilbereichen der Naturwissenschaften zu rationalisieren. Für die empiristisch orientierte Diagnostik in der Schule, in der die methodischen Regeln wissenschaftlichen Vorgehens im Kontext der Entsubjektivierungsforderung selbst als Paradigma fungieren, wird die Entkräftung des Vorwurfs Schwierigkeiten bereiten. Ihre Unmöglichkeit wird sich vor allem dort

offen zeigen, wo die Einführung von methodischen Vorgehensweisen als angeblich erkenntnisfunktional auf tautologischem Wege erfolgt. Es scheint sich daher in der empiristisch orientierten Diagnostik eine Schließung der "Rationalitätslücke" ⁷⁾ in Form einer verbindlichen, normativen Methodologie durchgesetzt zu haben, die zwar an die Konzeption von LAKATOS erinnert, aber a u s s c h l i e ß l i c h normativen Charakter hat und in ihrer Gegenstandsbezogenheit vollständig uninterpretiert ist. Dies steht im Gegensatz zur Funktionsbestimmung der Regeln der normalen Wissenschaft, wie sie im wissenschaftstheoretischen Rekonstruktionsversuch durch STEGMÜLLER vorgenommen wird, der dann jedenfalls als Vergleichsmaßstab heranzuziehen ist, wenn man sich als Vertreter der empiristisch orientierten Diagnostik nicht dem Irrationalismusvorwurf aussetzen will.

Auch wenn STEGMÜLLER, seiner wissenschaftstheoretischen Herkunft entsprechend, den normativen Einfluß von Methodologien und Methoden auf den Forschungsprozeß nicht eigens thematisiert, so leugnet er nicht die Tatsache, daß "jede wissenschaftstheoretische rationale Rekonstruktion nicht nur einen deskriptiven sondern einen normativen Aspekt besitzt und daß bisweilen der letztere stark in den Vordergrund treten kann. Trotzdem wäre es außerordentlich irreführend zu sagen, der Wissenschaftstheoretiker bemühe sich um die Gewinnung von Normen für rationales wissenschaftliches Verhalten" (STEGMÜLLER, Theoreindynamik, S. 299). Er bezeichnet deshalb den Versuch von LAKATOS, normative methodische Regeln aufstellen zu wollen, deren Befolgung eine abhanden gekommene Rationalität im Wissenschaftsprozeß wiederherstellen soll, als überflüssig und führt ihn - insoweit als

7) Der Begriff "Rationalitätslücke" beinhaltet eine der wichtigsten Kritiken an KUHNs Behauptungen über wissenschaftliche Revolutionen. Er soll die KUHN unterstellte Auffassung kennzeichnen, es gäbe in der Wissenschaftsentwicklung keinen kontinuierlichen und rekonstruierbaren Erkenntnisfortschritt. Dabei spielt u.a. die Frage eine Rolle, welche Funktion wissenschaftsmethodischen Regeln im Forschungsprozeß zukommt und ob ihre Befolgung für den Erkenntnisfortschritt konstitutiv ist (vgl. STEGMÜLLER, Theoriendynamik, S. 177f).

LAKATOS gemeint ist - auf die Fehldeutung des KUHNschen Begriffes von Irrationalismus zurück.

Darüberhinaus käme ein Bedürfnis nach der methodischen Reglementierung des Forschungsprozesses immer erst dann auf, wenn die betreffende Wissenschaft in einer Krise steckt, d.h., wenn sich die Anomalien häufen. Die Bewältigung dieser Anomalien sei aber keinesfalls ein methodisches Problem und schon garnicht durch eine methodische Reglementierung des Forschungsprozesses zu erreichen. Unter der Voraussetzung seines rationalen Rekonstruktionsversuches der Position KUHNs ist für ihn die Anomalienbewältigung ein Prozeß, an dem sowohl die normale Wissenschaft wie auch die außerordentliche Forschung partizipiert und der sich durch allgemeine Regeln wissenschaftlicher Tätigkeit nicht kennzeichnen läßt. Dies gilt auch deshalb, weil Wissenschaftler, Gesetze, Begriffe und Theorien nicht in abstracto lernen, sondern vielmehr nur in konkreten Anwendungen. Dabei kann generell gesagt werden: "Unter die Aktivitäten des normalen Wissenschaftlers ist alles zu subsumieren, was entweder mit einer Änderung von Kernerweiterungen bei gleichbleibenden Strukturkernen oder mit Änderungen in der Menge der intendierten Anwendungen zu tun hat" (STEGMÜLLER, Theoriendynamik, S. 183). Die außerordentliche Forschung ist dagegen mit der Errichtung neuer Strukturkerne beschäftigt, wobei sich die Entscheidung über die Brauchbarkeit einer Theorie nach o b j e k t i v e n E r f o l g s k r i t e r i e n bemißt, "die sich aus der Beantwortung der Frage ergeben, ob der neue Strukturkern erfolgreich für Kernerweiterungen benutzt werden konnte" (STEGMÜLLER, Theoriendynamik, S. 298).

Damit ist natürlich über eine möglicherweise positivierbare Bestimmung der Bedeutung von Reglementierungen im Forschungsprozeß nicht viel ausgesagt. Die Selbstverständlichkeit, mit der sich STEGMÜLLER gegen eine normative Funktionalisierung von Wissenschaftsregeln wendet und ihre Bedeutung jeweils in Abhängigkeit vom Stand der Entwicklung der gesamten Theorie sieht ("Der normale Wissenschaftler hält nicht kritiklos an Hypothe-

sen fest, sondern benützt ein begriffliches Instrument, dessen Beibehaltung mit wechselnden hypothetischen Annahmen verträglich ist ..." (ebd., S. 298)), mag verblüffen, wenn man an seine früheren wissenschaftstheoretischen Formalisierungsbestrebungen denkt. Sie ist wohl aber eine Konsequenz seiner gewandelten Auffassung in Richtung auf die "Dynamisierbarkeit" wissenschaftstheoretischer Konzeptionen und der Analyse des zugrundegelegten Gegenstandsbereiches mathematisierbarer Naturwissenschaften.

Angesichts der Tatsache, daß die Forschung in diesen Bereichen nach rekonstruierbaren, methodologischen Prinzipien und methodischen Regeln vorgeht, wäre es dringend erforderlich, diese Form der partiellen Rationalisierung von Forschungsparadigmen aus den Besonderheiten des Gegenstandsbereiches und dessen historisch-gesellschaftlicher Entwicklung zu begreifen. Dies umso mehr, als STEGMÜLLER selbst die Legitimität von wissenschaftlichen Vorgehensregeln, z.B. bestimmte Formen der Hypothesenbestätigung und Verfahren zur Einführung von theoretischen Begriffen in die Wissenschaftssprache sowie ihre Interpretierbarkeit, mit pragmatischen Zweckmäßigkeitsüberlegungen in Zusammenhang bringt. Die Behandlung dieser Frage wäre für das Übertragungsproblem in den empiristisch orientierten Sozialwissenschaften von besonderer Wichtigkeit, da durch einen Teilbereich der Regeln im wissenschaftlichen Vorgehen der Bezug zwischen Theorie und Empirie gestiftet werden soll, im Kontext des naiv-empiristischen Entsubjektivierungsparadigmas jedoch solche Regeln das Instrument der Gegenstandsanalyse bilden. Dort wird also das naturwissenschaftliche Vorgehen auf den Kopf gestellt.

Vor dem Hintergrund der Diskussion des KUHNschen Paradigmbegriffes läßt sich die Kennzeichnung der empirisch orientierten Diagnostik in der Schule in ihrer gegenstandsunbezogenen Ausrichtung an der Entsubjektivierungsforderung vorwegnehmend vollständiger durchführen. In Ermangelung einer historisch gewordenen Gegenstandsanalyse, wie sie sich in den exakten Naturwissenschaften entwickelt hat, und des sich darin vollziehenden, sukzessi-

ven Entsubjektivierungsprozesses, verfügt man in der empiristisch orientierten Diagnostik über kein vergleichbares Inventar von bewährten Prinzipien und Regeln wissenschaftlichen Vorgehens. Gleichzeitig hat sich aber im Zuge der zunehmenden Vergesellschaftung auch der Sozialwissenschaften der Entsubjektivierungsanspruch in verselbständigter Form durchgesetzt und muß nun als erfolgsversprechendes Paradigma im KUHNschen Sinn durch die normale Wissenschaft eingelöst werden. Da eine eigene erfolgreiche Forschungstradition fehlt, erscheint die Anlehnung an die empiristische Wissenschaftstheorie und an das Vorgehen der exakten Naturwissenschaften nicht nur plausibel, sondern notwendig.

Daraus ergibt sich eine fatale Konsequenz: Da das Entsubjektivierungsparadigma nicht entsprechend dem STEGMÜLLERSchen Rekonstruktionsversuch rationalisiert und interpretiert werden kann - aufgrund der fehlenden Gegenstandsanalyse ist die Systematisierung des wissenschaftlichen Erkenntnisstandes nach Strukturkern, nach durch paradigmatische Beispiele festgelegten Anwendungsbereich, nach Bereich der Kernerweiterung usw. nicht möglich -, wäre eine Kommunikation zwischen Wissenschaftlern auch innerhalb der empiristischen Richtung nicht zu erwarten. Andererseits kann man aber eine Kommunikationsaktivität feststellen, die den Anschein des "Sich-Verstehens" erweckt. Es besteht aber der zu begründende Verdacht, daß sich die gegenseitige Verständigung auf die Vermittlung von Regeln normalwissenschaftlicher Tätigkeit beschränkt. Deswegen ist man wohl immer noch sehr stark an einer Trennung zwischen Meßtheorie und Testtheorie von der angewandten Psychologie interessiert (FISCHER, 1974).

Die Aufrechterhaltung dieser Form von Verständigungsmöglichkeit, die um den Preis einer strikten Trennung von Inhalt und Methode erkaufte werden muß, läßt den Schluß zu, daß die Wirksamkeit des Entsubjektivierungsparadigmas in der Verwissenschaftlichung der Diagnostik durch die KUHNsche Auffassung in der Interpretation von POPPER beschrieben werden kann. Sie erklärt außerdem, warum

eine bestehende Verständigung zwischen Beurteilern immer nur ex post facto zu konstatieren ist - "Objektivität liegt dann vor, wenn die Beurteiler übereinstimmen". Und sie bestätigt schließlich die Wirksamkeit des Entsubjektivierungsparadigmas im diagnostischen Forschungsprozeß in der Funktion einer normativen Methodologie, wie sie von LAKATOS gefordert wurde. Diesbezüglich sei noch einmal an WESTMEYER erinnert: "Es gehört zum Wesen eines konstruktivistischen Ansatzes, daß er sich nicht lange bei der Kritik des Gegebenen aufhält, sondern Normen einzuführen versucht, nach denen hier und heute gehandelt werden soll ...".

1.5 Zusammenfassung und Problemstellung

Insgesamt wird mit der Arbeit das Ziel verfolgt, Aussagen über Schüler in Form von Urteilen über deren Lernverhalten auf ihren Bedeutungsgehalt zu untersuchen. Damit steht zugleich die Frage im Hintergrund, ob derartige Aussagen im Verwendungskontext des bestehenden Schulsystems intersubjektiv konsistent interpretiert werden können. Deshalb war es aus Gründen der Systematik notwendig, die "Objektivitätsproblematik" an den Anfang der Untersuchung zu stellen. Den äußeren Rahmen für die systematische Priorisierung der Objektivitätsproblematik bildeten drei feststellbare Tatbestände:

- die seit langem bekannte und reflektierte Kritik am Subjektivismus des Lehrerurteils,
- die langsam das öffentliche Bewußtsein prägende Forderung nach einer Objektivierung bzw. Entsubjektivierung des Schülerbeurteilungsprozesses und
- der Anspruch der wissenschaftlichen, pädagogisch-psychologischen Diagnostik, die Entsubjektivierungsforderung einzulösen.

Von entscheidender Bedeutung ist der Entsubjektivierungsanspruch, wie er von den empiristisch orientierten Wissenschaftlern innerhalb der diagnostischen Forschung vertreten wird. Er nimmt im Forschungsprozeß einen programmatischen Stellenwert ein. Da sich die damit verbundenen Entsubjektivierungsvorhaben des diagnostischen Prozesses einerseits auf das konstatierbare Dilemma im Bereich der Schülerbeurteilung beziehen und andererseits ihre Legitimation und die ihnen unterstellte Erfolgsaussicht aus der Orientierung der diagnostischen Forschung an der empiristischen Wissenschaftstheorie gewonnen werden, waren die Erscheinungsformen des Objektivitätsproblems systematisch zusammengefaßt darzustellen; sie waren mit jenen Bedingungen zu konfrontieren, die wissenschaftstheoretisch an den intersubjektiven Gebrauch von wissenschaftlichen Aussagen gestellt werden.

Die Untersuchung zeigt, daß man sich im Vollzug der Verwissenschaftlichung der Diagnostik in der Schule lediglich an einer

Bedingung orientiert: an der Einlösung der Objektivitätsforderung im Sinne einer formalisierten Intersubjektivität, wie sie der Struktur des Objektivitätsbegriffes in der klassischen Testtheorie entspricht. Die beiden anderen Bedingungen, die im Rahmen eines empiristischen Wissenschaftsmodells für entsubjektivierte, wissenschaftliche Aussagen gelten müssen - die Intersubjektivität im Gebrauch einer Beobachtungssprache und in der Interpretation von als invariant betrachteten Verhaltensmustern - bleiben weitgehend unberücksichtigt.

Dies führt dazu, daß man mit Hilfe methodisch objektivierter Schülerbeurteilungen Resultate erhält, die für den **k o n - k r e t e n E i n z e l f a l l** die Aussagefähigkeit der Zensuren nicht übersteigen und dieses Phänomen selbst in Ermangelung einer diagnostischen Theorie nicht erklärt werden kann. Die Gründe dafür sind in der vom Standpunkt der empiristischen Wissenschaftstheorie fragwürdigen Begrifflichkeit diagnostischer Aussagen und in der tautologischen Struktur diagnostischer Hypothesenbildung zu suchen. Das Vorgehen bei der Verwissenschaftlichung der Diagnostik in der Schule läßt sich als naiv-empiristisches dem eigentlich intendierten logisch-empiristischen Vorgehen gegenüberstellen und sich dadurch kennzeichnen, daß die damit gewonnenen, methodisch objektivierten Beurteilungsergebnisse ebensowenig intersubjektiv konsistent zu interpretieren sind wie die kritisierten Urteile in Form von Zensuren. Zugleich eröffnet sich bei der Untersuchung der Verwissenschaftlichungsprozesse im Bereich der schulischen Diagnostik eine Betrachtungsperspektive, unter der die Bestimmung der Funktion der Regeln zur Objektivierung der Beurteilung erforderlich wird.

Aufgrund der Diskrepanz zwischen dem Anspruch der empiristisch orientierten Diagnostik in der Schule und dem Ergebnis der Anwendung objektivierter Verfahren zur Schülerbeurteilung kann man vermuten, daß der verwendete Methodenkanon im Forschungsprozeß keine erkenntnisfunktionale Bedeutung hat, sondern als normative Methodologie fungiert, wie dies aus dem Konzept des "Paradigmas" von KUHIN ableitbar zu sein scheint. Die Art und

Weise des Umgangs mit den formal-methodischen Regeln wissenschaftlicher Tätigkeit und deren sukzessive Verselbständigung im Bereich der Diagnostik ist mit jener kriseninduzierenden Forschungsphase vergleichbar, in der sich nach KUHN die Anomalien häufen und mit den tradierten Vorstellungen nicht mehr zu bewältigen sind. Hervortretendes Merkmal der Verselbständigung ist die gegenstandsunbezogene Anwendung und Vermittlung des Methodenkanons, d.h., die Reglementierung der Forschungstätigkeit durch methodische Prinzipien und Verfahren zur Entsubjektivierung der Hypothesengewinnung und -überprüfung ohne Bezug zu einer differenzierten Theorie oder zur konkreten Erforschung des Lernverhaltens in der Durchführung und im Verlauf des organisierten Lernens.

Die These der Verselbständigung der methodischen Objektivierung diagnostischer Prozesse in der Schule im Sinne einer normativen Methodologie steht im Hintergrund der gesamten Arbeit und soll im weiteren detailliert ausgeführt werden. Aus diesem Grund war vorher eine Auseinandersetzung mit dem "Paradigma"-Konzept im allgemeinen durchzuführen und die Entstehung und Funktion des Entsubjektivierungsparadigmas im Bereich der Diagnostik in der Schule zu behandeln. Die Untersuchung des letzteren ging aus von der Gegenüberstellung zweier exemplarischer Funktionen des Entsubjektivierungsparadigmas. Als wichtigstes Ergebnis der Gegenüberstellung ist festzuhalten, daß, bezogen auf den Forschungsprozeß der exakten Naturwissenschaften, die Formulierbarkeit eines Forschungsparadigmas die Konsequenz tradierter wissenschaftlicher Tätigkeit darstellt, während sie in der Diagnostik normative Vorbedingung ist. Von da aus betrachtet waren die Bedingungen zu analysieren, unter denen sich das Entsubjektivierungsparadigma im Prozeß der Verwissenschaftlichung der Diagnostik in der Schule in seiner normativen Funktion durchsetzen konnte. Die Diskussion seiner geschichtlichen Entstehung bzw. Durchsetzung führt auf drei kategoriale Bedingungen:

1. seine psychotechnisch-diagnostische "Bewährungsprobe" in außerschulischen Bereichen,
2. seine Objektivierungsleistung in Bezug auf die Kritik am Lehrerurteil und
3. seine Funktionalität für die Technologisierung gesellschaftlicher Prozesse

Die Untersuchungen der Bedingungen für den Durchbruch des Entsubjektivierungsparadigmas liefern gleichzeitig das allgemeine Kennzeichen einer bestimmten Funktion von Paradigmen im Forschungsprozeß - die von KUHN beschriebene Steuerung der wissenschaftlichen Tätigkeit durch das In-Aussicht-Stellen von wissenschaftlichen Erfolgen, obwohl die einzelnen sog. normalen Wissenschaftler in einem "irrationalen" Verhältnis zu diesem Paradigma stehen.

Die Auseinandersetzung mit der KUHNschen Auffassung zur Wissenschaftsentwicklung unter Bezug auf den "rationalen" Rekonstruktionsversuch von STEGMÜLLER mündet in zwei Thesen, deren Gültigkeit für die Verwissenschaftlichung der Diagnostik in der Schule zu überprüfen sind:

1. Die Ansätze zur methodischen Objektivierung der Schülerbeurteilung basieren auf formalen Strukturen, für die keine definierten Anwendungsbereiche existieren.
2. Der intersubjektiv konsistente Gebrauch von diagnostischen Aussagen innerhalb der Wissenschaftlergemeinschaft und/oder der Beurteilergemeinschaft beschränkt sich auf deren methodische Konstitution bzw. auf deren verfahrenstechnische Konstruktion und Anwendung.

In der Diskussion der ersten These wird es um die Untersuchung der empiristisch wissenschaftstheoretischen Voraussetzungen für die Gewinnung von diagnostischen Aussagen gehen. Dabei wird zu fragen sein, ob durch die Verfahrensweise der Lernzielbestimmung eine brauchbare Grundlage für die Konzipierung von Urteilsaussagen geschaffen wird, ob die diagnostische Begriffsbildung empirisch zu stützen ist und ob die praktizierte Formulierung von "Erfahrungshypothesen" zu einem besseren Verständnis des Lernverhaltens führen kann.

Bei der zweiten These steht der Prozeß der praktizierten Diagnostik in der Schule im Mittelpunkt der Betrachtung und zwar sowohl die Zensurengebung als auch die Bemühungen um eine Objektivierung der Schülerbeurteilung. Es wird zu zeigen sein, daß sich die Kritik am Subjektivismus des Lehrerurteils durch die Objektivierungstendenzen nicht aufheben läßt, sondern nur auf die Interpretation der Beurteilungsergebnisse verschoben wird.

2. Der Gegenstandsbereich der Schülerbeurteilung

Die Untersuchung beider Thesen, die am Ende der Zusammenfassung formuliert wurden, erfolgt im weiteren Verlauf der Arbeit im Zusammenhang. An den Nachweis, daß die Ansätze zur methodischen Objektivierung der Schülerbeurteilung auf formalen Strukturen basieren, für die keine definierten Anwendungsbereiche existieren, muß sich bei der Behandlung ausgewählter diagnostischer Fragestellungen jeweils die Problematisierung der Verständigungsbasis der Wissenschaftler- bzw. Beurteilergemeinschaft anschließen. Die Behauptung, daß sich der intersubjektiv konsistente Gebrauch von diagnostischen Aussagen innerhalb dieser Gemeinschaften auf die methodische Konstitution von Beurteilungsverfahren beschränkt, wird sich dann als richtig erweisen, wenn innerhalb der Wissenschaftler- bzw. Beurteilergemeinschaft keine Einigung in der Festlegung eines verbindlichen Anwendungsbereiches einer Beurteilungsmethode zu finden ist. Deshalb steht die Betrachtung der ersten These im Vordergrund.

Dazu ist allerdings von einer weitergehenden Präzisierung auszugehen. Genau genommen, wird in dieser These nicht behauptet, daß für die Diagnostik in der Schule kein Anwendungsbereich bestimmbar wäre; der Anwendungsbereich ist durch die objektiven Bedingungen determiniert, die der Schülerbeurteilung durch das selektive Ausbildungssystem vorgegeben sind und die allgemein die Diagnose von prognostisch relevanten, auf die Ausbildungsalternativen bezogenen Schülereigenschaften erforderlich machen. Für genau diese Aufgabe scheinen sich allerdings die der methodischen Objektivierung zugrundeliegenden formalen Strukturen des Beurteilungsprozesses nicht zu eignen. Deshalb geht es bei der Untersuchung der ersten These um die Überprüfung der Behauptung, daß die Beziehung zwischen dem noch näher zu explizierenden Anwendungsbereich und der formalen Strukturierung der Schülerbeurteilung nicht in der Weise zu "rationalisieren" ist, wie dies durch den wissenschaftstheoretischen Rekonstruktions-

versuch bezüglich der KUHNschen Position erforderlich wird. Die Überprüfung der so präzisierten Behauptung begegnet allerdings zwei Arten von Schwierigkeiten:

1. Der rationale Rekonstruktionsversuch der KUHNschen Position ist ein logisch-mathematisch hochformalisiertes Verfahren, stellt enorme Ansprüche an die Präzision der Begriffseinführung und -verwendung und hat als gegenständliche Entwicklungsgrundlage die theoretische Physik, deren Mathematisiertheit ja erst die Voraussetzung schuf, den Prozeß der Theorienbildung und -entwicklung auf der Ebene der Unterscheidung von Strukturkern einer Theorie und ihrem intendierten Anwendungsbereich darzustellen. Wollte man den Rekonstruktionsversuch in seiner gesamten Komplexität auf den sozialwissenschaftlichen Forschungsprozeß übertragen, so müßte man die Mathematisiertheit des jeweiligen sozialwissenschaftlichen Gegenstandsbereiches voraussetzen können. Zudem kann man das Übertragungsproblem durch jene Bedenken charakterisieren, die DIEDERICH gegenüber dem Rekonstruktionsversuch äußert:

"Erstens ist noch offen, wie weit sich Sneeds Analyse der Theorienstruktur in der mathematischen Physik über diesen Wissenschaftszweig hinaus erweitern oder sinngemäß übertragen läßt; Kuhn hat ja das Entwicklungsmuster aller Naturwissenschaften, also auch weniger quantifizierter und formal durchgebildeter (Sperrung vom Verfasser), als die theoretische Physik es ist, beschreiben wollen. Zweitens fragt sich, ob einige wesentliche Konzepte Kuhns, selbst wenn man sie auf die mathematische Physik restringiert, adäquat rekonstruiert sind" (id., 1974, S. 34).

Andererseits wird jedoch die Entwicklung von methodisch objektivierten Formen der Schülerbeurteilung durch den Anspruch begründet, ein Maximum an Entsubjektivierung solcher Urteilsaussagen zu erreichen. Und empiristisch wissenschaftstheoretisch betrachtet, ist die Entsubjektivierung des Forschungsprozesses in Wissenschaftsdisziplinen am weitesten fortgeschritten, in denen man fast nur noch mit quantitativen Begriffen arbeitet. Deshalb ist der wissenschaftstheoretische Rekonstruktionsversuch ein optimaler Maßstab für die Beurteilung des Stellenwertes der Entsubjektivierungsregeln im diagnostischen Forschungsprozeß.

Wegen der fehlenden Voraussetzungen in Bezug auf die Präzisionsanforderungen an die wissenschaftliche Reflexionstätigkeit muß sich aber die Untersuchung der These auf eine weitgehend "intuitive" Verwendung des wissenschaftstheoretischen Maßstabes beschränken.

2. Damit zusammen hängt das Problem, daß in der verwissenschaftlichten Diagnostik in der Schule zwar formalisierte Kalküle zur Anwendung kommen (z.B. verschiedene Arten von Testtheorien), diese aber nicht als Strukturkerne im Sinne einer mathematischen Struktur, welche allen diagnostischen Gegenstandsbereichen eigentümlich ist, interpretiert werden können. Im Gegensatz dazu müssen sie sich erst als pragmatisch brauchbare Formen der Darstellung von diagnostischen Prozessen bewähren.

Dies hat zur Folge, daß für die Untersuchung der These ein entgegengesetzter Weg einzuschlagen ist: man kann nicht prüfen, ob umschreibbare Anwendungsbereiche die Bedingungen einer formalisierten Struktur erfüllen und also Modelle derselben darstellen, sondern ob die formalisierten Strukturen den Anforderungen entsprechen, die im Anwendungsbereich vor aller Verwissenschaftlichung festliegen, z.B., ob für den Anwendungsbereich "selektionsorientierte Schülerbeurteilung" die klassische Testtheorie zu intersubjektiv konsistent interpretierbaren Aussagen über Dispositionen von Schülern führt. Erst dadurch wird sich die "Irrationalität" der Beziehung zwischen formaler Struktur und Anwendungsbereich zeigen.

Für den Anwendungsbereich der Beurteilung von dispositionellen Schülereigenschaften mit dem Anspruch der prognostischen Relevanz der Urteile muß deshalb die Bedeutung von Dispositionsbegriffen diskutiert und ihre Verwendung in diagnostischen Aussagen untersucht werden, die die Prognose des Lernerfolges von Schülern zum Ziel haben.

2.1 Empirische Bedeutung und empiristische Rekonstruktion von Dispositionen

Zur Entwicklung der Problematik von Dispositionsbegriffen gehe ich von einer allgemeinen Klärung der selektionsbezogenen Diagnostik in der Schule aus. Es wird sich dabei eine Umschreibung ihres Anwendungsbereiches ergeben, für den sich die Bezeichnung "Verwendungsbereich" besser eignen würde. Es handelt sich um die Bestimmung eines Gegenstandsbereiches, der nicht potentiell ein Modell für eine theoriekernartige, mathematische Struktur sein könnte, sondern der unabhängig von und vor seiner wissenschaftlichen Durchdringung durch gesellschaftliche Bedingungen des Ausbildungssystems relativ restringiert vorstrukturiert ist. Ich werde aber, solange die wissenschaftstheoretische Betrachtung im Hintergrund steht, den Terminus "Anwendungsbereich" benutzen und erst unter Rekurs auf den Selektionsaspekt von Verwendungsbereich sprechen.

Das zentrale Problem der pädagogisch-psychologischen Diagnostik besteht m.E. darin, daß die zu diagnostizierenden Dispositionen von Individuen in einer spezifischen Weise der unmittelbaren Beurteilung nicht zugänglich sind, unabhängig davon, ob durch den Beurteilungsprozeß eine Klassifikation der Dispositionen, eine Einschätzung auf dem Niveau einer Ordinalskala oder eine Messung vorgenommen werden soll. Grundlage für die Beurteilung sind die verschiedenen Vergegenständlichungsformen des Schülers im Rahmen des organisierten Lernens als Resultat seiner Auseinandersetzung mit den komplexen Situationen "Unterricht" bzw. "Schule" und mit deren Veränderung in der Zeit; Grundlage ist aber nicht der empirische Auseinandersetzungsprozeß selbst.

In wesentlichen Teilen ist daher die verwissenschaftlichte, pädagogisch-psychologische Diagnostik als Spezialdisziplin der psychologischen Methodik an jene methodischen Restriktionen gebunden, die die empirische Sozialwissenschaft allgemein charak-

terisieren:

"Im Gegenstandsgebiet der Psychologie ist nur das Verhalten bestimmter Subjekte in bestimmten Situationen dem methodischen Zugriff unmittelbar zugänglich. Nur zwischen den Elementen aus der Menge von Verhaltensweisen, verschiedenen Subjekten und verschiedenen Situationen und der kartesischen Produkte dieser Ausgangsmengen lassen sich empirische Relationen aufzeigen. Solche Befunde werden als Daten bezeichnet und sind z.B. Leistungen verschieden intelligenter Personen bei verschiedenen Testitems oder die verbalisierten Wahrnehmungseindrücke verschiedener Vpn unter verschiedenen Reizbedingungen. Was man in der Psychologie aber - ihrer praktischen Aufgabenstellung und ihrer theoriebildenden Funktion gemäß - messen will, sind z.B. die Intelligenz, Einstellungen, phänomenale Ähnlichkeiten und Motivationen usw. Das Verhalten wird dabei als Indikator benutzt und dient vornehmlich dazu, an die direkt nicht zugänglichen Eigenschaften der Subjekte (als Verhaltensweisen) und ihrer Wahrnehmungseindrücke (als Reizwirkungen) heranzukommen und diese (Indikanden) der Messung zugänglich zu machen" (KUKLA, 1973, S. 117).

Aus der Tatsache, das in die Beurteilung von Dispositionen das "Bestimmte" der das Verhalten provozierenden Situationen und das "Bestimmte" des als provoziert betrachteten Verhaltens eingeht, resultiert für die wissenschaftliche Untersuchung des Beurteilungsprozesses die Notwendigkeit, das Beurteilungsergebnis unter zwei Gesichtspunkten erkenntniskritisch zu beleuchten:

1. Im Hinblick auf welche gesellschaftliche Bedeutung und theoretische Projektion läßt sich die Bestimmung von Situationen und Verhaltensweisen als Bedingung für die Aktivierung von daraus zu erschließenden Dispositionen begründen?
2. Inwieweit kommt solchen Dispositionen empirische Objektivität zu, d.h., inwieweit sind die Verallgemeinerungen von materiellen Prozessen der Auseinandersetzung des Individuums mit den provozierenden Situationen und nicht nur als Konstrukte in Form rein verbaler, theoretischer, erfundener, subjektiv konstituierter Gebilde faßbar, die nur der Wirklichkeitssphäre einer theoretischen Wissenschaftssprache angehören? 8)

8) Die kritische Entgegensetzung der Forderung nach empirischer Objektivität des Dispositionsbegriffes gegenüber seiner Be-

Beide Fragen stellen sich im Kontext pädagogisch-psychologischer Diagnostik in Bezug auf die besondere Funktion von Dispositionsbegriffen zur Erklärung von Reiz-Reaktionskontiguitäten. In *Analogue* zu den Naturwissenschaften soll mit Dispositionsbegriffen die Potentialität von Gegenständen zum Ausdruck gebracht werden, sich unter angebbaren Bedingungen in angebbare Zustände zu transformieren. (Beispiele seien die bekannten Dispositionsbegriffe "Wasserlöslichkeit" und "Zerbrechlichkeit"). Die Probleme, die bei der Einführung von Dispositionsbegriffen in die Wissenschaftssprache auftragen, bezogen sich auf die Einlösung des Anspruchs, diese Begriffe vollständig empirisch interpretieren zu können. Dabei ging es aber nicht mehr um die eindeutige und präzise Bestimmung der jeweiligen Bedingungen, Transformationen und Gegenstände als Grundlage für die Einführung der Begriffe, sondern um das Finden einer allgemeingültigen logischen Form der Reduktion der Dispositionsbegriffe auf Begriffe und Aussagen der Beobachtungssprache, deren empirische Gedeutetheit vorausgesetzt wurde.

Die zunächst herangezogene Reduktionsform der operationalen Definition traf sowohl auf einen logischen Einwand als auch auf einen Einwand, der die empirische Deutbarkeit des Dispositionsbegriffes betraf. Der in der operationalen Definition enthaltene Allquantor ließ keine empirische Verifikation der darin ebenfalls formulierten Hypothese zu, während die Spezifität der definierenden Operationen die vollständige, empirische Interpretation des Dispositionsbegriffes verhinderte.

deutung als ausschließlich wissenschaftliches Gebilde ist nicht nur vom Standpunkt einer materialistischen Empirie aus gerechtfertigt. Auch aus der Perspektive der "dynamisierbaren" Wissenschaftstheorie bleibt die Gegenüberstellung in einem eingeschränkten Sinn bedeutungsvoll, obwohl STEGMÜLLER die mathematische Struktur einer Theorie als *reine begriffliche Gebilde* bezeichnet. Man muß dabei allerdings die für STEGMÜLLER in Bezug auf die Theoriebildung gegenstandslos gewordene Unterscheidung von Entdeckungszusammenhang und Rechtfertigungszusammenhang durchführen und in einer historischen Untersuchung aufzeigen, daß dieses heute "rational" rekonstruierbare, rein begriffliche

Auf die einzelnen, gescheiterten Versuche der Beseitigung des Dilemmas kommt es hier nicht an; wesentlich ist das Resultat der Entwicklung, die Einführung der Dispositionsbegriffe als theoretische in die Wissenschaftssprache (vgl. STEGMÜLLER, 1970, Band II, Studienausgabe Teil B). Für sie gilt nicht mehr die BRIDGEMANsche Forderung der vollständigen Rückführbarkeit der Dispositionsbegriffe auf Begriffe der Beobachtungssprache durch operationale Definitionen. Für Dispositionsbegriffe wie auch, in anderem Zusammenhang, für metrische Begriffe ist ein inhaltlicher Bedeutungsüberschuß zugelassen, der als Hypothesenquelle fungiert und die Theorienbildung vorantreibt.

In Verbindung mit den beiden vorher gestellten Fragen und der noch zu behandelnden, analogen Verwendung von Dispositionsbegriffen in der Diagnostik in der Schule ist das Phänomen zu berücksichtigen, daß die Problematisierung der Dispositionsbegriffe im modernen Empirismus nur die Forderung nach empirischer Objektivität berührt und sich dabei wiederum nur auf einen eingeschränkten Aspekt bezieht, nämlich die methodologische Absicherung der Beziehung zwischen theoretischer Sprache und Beobachtungssprache. STEGMÜLLER präzisiert den eingeschränkten As-

"Gebilde" (STEGMÜLLER, 1973, Band II, Studienausgabe Teil D, S. 122). im Prozeß konkreter, naturwissenschaftlicher Forschung sukzessiv entstanden ist, und kein empirieloses, subjektivistisch konstituiertes Gebilde im Sinne der zweiten Frage darstellt. Dies historisch nachzuweisen wäre gerade am Beispiel der klassischen Partikelmechanik möglich, wo durch das Prädikat "ist eine klassische Partikelmechanik" die thematische Struktur einer Theorie charakterisiert wird (vgl., ebd., S. 107ff).

Im Gegensatz dazu dürfte historisch gesehen die Entstehung der Testtheorie als Regelkomplex für die Durchführung diagnostischer Tätigkeiten weitgehend unbeeinflusst von den psychologischen Analysen der Verarbeitung von diagnostischen Bedingungen gewesen sein. So hat sich z.B. der oft als spiritus rector der Testologie zitierte GALTON letztlich nicht dafür interessiert: "... er wollte quantitativ ein Resultat festhalten und nicht die physiologischen und psychologischen Bedingungen aufdecken, die hinter den Leistungen stecken ... In diesem Zusammenhang darf man seinen darwinistischen Standort nicht übersehen, der ihn am Problem der Variabilität überhaupt mehr als an der Intelligenz alleine interessiert sein ließ" (GROFFMANN, 1964, S. 158).

pekt in der Weise, daß er auf die Konfusion hinweist, "welche die Diskussion über die theoretischen Begriffe immer wieder unnötig aufgehalten hat, nämlich die mangelnde Unterscheidung zwischen dem Fehlen beobachtungsmäßiger Kriterien für das Vorliegen eines Begriffes und der Undefinierbarkeit dieses Begriffes innerhalb einer Beobachtungssprache" (STEGMÜLLER, 1970, Band II, Studienausgabe Teil A, S. 7).

Die gesellschaftliche Bedeutung der in der Wissenschaftssprache verwendeten Begriffe schien durch die Intersubjektivität ihres Gebrauchs als gegeben voraussetzbar zu sein: Für die Beobachtungssprache durch den bewährten Verwendungskonsensus in der Wissenschaftlergemeinschaft, für die sog. Zuordnungsregeln in ihrer Eigenschaft als partiell empirische Interpretationsregeln theoretischer Ausdrücke durch die ihnen unterstellte Zweckmäßigkeit für die Theorienbildung, für die nur partiell empirisch interpretierbaren theoretischen Begriffe durch den Konsensus in Bezug auf ihre sog. prognostische Relevanz und den in ihnen enthaltenen Bedeutungsüberschuß als Potenz weiterer Theorienbildung (vgl. zur Entwicklung des empiristischen Sinnkriteriums STEGMÜLLER, 1970, Band II, Studienausgabe Teil B und Teil C).⁹⁾.

-
- 9) Der hier verwendete Begriff der "prognostischen Relevanz" hat eine völlig andere Struktur als der häufig damit gleichgesetzte testtheoretische Begriff der "prognostischen Gültigkeit". Umgangssprachlich ausgedrückt hat ein bestimmter theoretischer Begriff dann prognostische Relevanz, wenn durch seine Hereinnahme in die Theorie Voraussagen beobachtbarer Ereignisse abgeleitet werden können, die bei seinem Ausschuß nicht möglich sind. Die Aufmerksamkeit kann bei dieser umgangssprachlichen Formulierung nicht stark genug auf das Wort "abgeleitet" gelenkt werden, um der möglichen Assoziation einer Zukunftsvorhersage von vornherein zu begegnen. Letzteres läßt sich für den Begriff der prognostischen Gültigkeit kaum vermeiden, da für die Erhöhung der die prognostische Gültigkeit indizierenden Korrelationskoeffizienten weder eine Gegenstandstheorie notwendig ist, noch die zugrundeliegenden Urteilsaussagen über Schüler abgeleitet werden können. Berücksichtigt man zudem das ungelöste Problem der Beobachtungssprache, die als intersubjektiv gebrauchte der Beschreibung des Schülerverhaltens dienen soll, dann erscheint die Bezeichnung der prognostischen Gültigkeit als empirische rätselhaft.

In Entsprechung zum ausschließlich wissenschaftsmethodologischen Bewältigungsversuch des Problems der empirischen Objektivität, bleibt der materiale Konstitutionsprozeß des Gegenstandes und seiner Veränderungen als Ausdruck gesellschaftlich-historischer Möglichkeiten außerhalb des Reflexionshorizontes. Er liefert aber dennoch in seiner Allgemeinheit die Grundlage für den intersubjektiven Gebrauch einer Beobachtungssprache. Deshalb erscheint die als gesellschaftlich verallgemeinerbare und durchgesetzte Bedeutung der wissenschaftlichen Begrifflichkeit in ihrer Eigenschaft als gesellschaftlich allgemeine nur noch als intersubjektiv gebrauchte. Die Frage, ob die an den Gegenstandsbereichen der Naturwissenschaften vor dem und im Forschungsprozeß vollzogenen empirischen Operationen ein Ausmaß der Verallgemeinerung im Verlauf ihrer geschichtlichen Entwicklung erreicht haben, das die Reflexion der gesellschaftlichen Bedeutung der diese empirischen Operationen abbildenden Begrifflichkeit als überflüssig erscheinen läßt - und gerade deswegen das Postulat des intersubjektiven Gebrauchs der Wissenschaftssprache ermöglicht -, wird nicht mehr gestellt.

Gleichwohl bleibt das Problem der historischen Rekonstruktion der gesellschaftlichen Bedeutung der naturwissenschaftlichen Begrifflichkeit als solches bestehen. Es tritt in der Behandlung durch den modernen Empirismus als zu formalisierendes dort besonders deutlich hervor, wo der Anspruch der Zurückführbarkeit nicht-analystisch sinnvoller Aussagen auf Aussagen der Beobachtungssprache bzw. theoretischer Begriffe auf die Beobachtungssprache preisgegeben werden muß. Exemplarisch wird dies bei der Einführung von klassifikatorischen, komparativen und metrischen Begriffen in die Wissenschaftssprache demonstriert. In den Einführungsregeln sind gleichzeitig Konventionen, Zweckmäßigkeitsüberlegungen wie Einfachheit und Fruchtbarkeit, als auch empirische Operationen miteinander verflochten. STEGMÜLLER legt hier ausdrücklichen Wert auf die Hervorhebung der Bedeutsamkeit dieser Bestimmungsmomente als Komponenten der Einführung von Begriffen.

Die Selbstverständlichkeit, mit der man in den Naturwissenschaften die Wirksamkeit normativer Bedingungen im Forschungsprozeß aufgrund des zunehmend entsubjektivierten, instrumentalisierten Vorgehens übersehen könnte, erweist sich heute als konkrete Schwierigkeit für die vollständige Rekonstruktion der Wissenschaftsgeschichte mit gesellschaftlichen Kategorien.

"Man kann zwar vielleicht den Zusammenhang zwischen bestimmten Schwerpunkten naturwissenschaftlicher Forschung und möglicherweise auch noch Paradigmata, Entwicklungsbedingungen und Wertungsinteressen des Kapitals herstellen. Was aber bisher noch nicht - bzw. in bescheidenen Ansätzen geleistet worden ist, ist die Rekonstruktion des Zusammenhanges von Wertabstraktion und Kategorien der Naturwissenschaft. Hierunter sind diejenigen Begriffe zu verstehen, ohne die z.B. eine Physik überhaupt nicht gedacht werden kann, für die klassische Physik z.B.: Raum, Zeit, Masse, Kraft. Unmittelbar ist diesen Begriffen auch nicht die Spur von Werthaftigkeit im Sinne des Einschließens von subjektiven oder auch gesellschaftlich bedingten Interessen anzusehen" (BRINKMANN, u.a., 1973, S. 28).

Das Scheitern des Empirismusprogramms provozierte in seiner Entwicklung die zunehmende Notwendigkeit, die Bedeutung von wissenschaftlichen Begriffen, die einerseits auf der analytischen Ebene nicht abgedeckt werden kann und andererseits aus der Beobachtungssprache nicht vollständig zu synthetisieren ist, durch die in die Wissenschaft transformierten, gesellschaftlichen Ansprüche in Form von Konventionen und Zweckmäßigkeitsüberlegungen mitfestzulegen und das gesamte Verfahren als wissenschaftlich zu legitimieren. Dabei konnte man auf die Begründung der Konventionen und Zweckmäßigkeitsüberlegungen scheinbar deshalb verzichten, weil die gesellschaftliche Allgemeinheit der Wissenschaftssprache durch den materiellen Forschungsprozeß in den Naturwissenschaften als Grundlage der Beobachtungssprache gegeben ist.

Durch den Verzicht auf die Begründung ist es allerdings nicht nur plausibel, sondern auch vielmehr erforderlich, mit dem Fortschritt der Naturwissenschaften die Konventionen und Zweckmäßigkeitsüberlegungen einer permanenten Veränderung zu unterziehen - was jedoch die Konsequenz nach sich zieht, den empirisch nicht interpretierbaren Rest der Begriffsbedeutung nur paradigmatisch

auf den historisch gegebenen Forschungsstand beziehen zu können. KUHN hat in seiner "Struktur wissenschaftlicher Revolutionen" diese Konsequenz programmatisch zu legitimieren versucht. Es wäre im einzelnen an der geschichtlichen Entwicklung der Naturwissenschaften aufzuzeigen, warum trotz der Bewußtlosigkeit der naturwissenschaftlichen Forscher gegenüber der gesellschaftlich-historischen Bedingtheit ihres Vorgehens ein eindeutiger und unbezweifelbarer Erkenntnisfortschritt zu verzeichnen ist. Die geschichtliche Rekonstruktion der Entwicklung der Naturwissenschaften läge vor allem im Interesse einer empiristisch orientierten Sozialforschung, die in anbetracht des Erfolges der naturwissenschaftlichen Forschung dort entwickelte methodologische Prinzipien und Methodiken formal übertragen zu können glaubt. Es würde sich dabei zeigen, daß die Entwicklung von Methodologie und Methodik das Ergebnis der unmittelbaren Auseinandersetzung mit dem Gegenstandsbereich ist und sich gerade dadurch die Notwendigkeit ergibt, bestimmte ideologische, längerfristig wirksame Erkenntnisparadigmen aufzugeben. KUHNs historisierende Zusammenfassung der Entwicklung der Naturwissenschaft bestätigt diese These ebenso wie die auf den Problemkreis "Erklären und Voraussagen" bezogene Betrachtung naturwissenschaftlicher Vorgehensweise durch TOULMIN (1968).

Wenn auch der wissenschaftsgeschichtliche Paradigmenwechsel im Zusammenhang mit dem Fortschreiten der sog. normalen Wissenschaft bei KUHN nicht ansatzweise auf seine gesellschaftlichen Bedingungen zurückgeführt wird, so verweist gerade die Tatsache des "normalen" Wissenschaftsfortschritt auf die Notwendigkeit, die Konstituiertheit der Gegenstandsbereiche, auf die sich die Naturwissenschaft jeweils beziehen konnte und die sie als Quasivorgegebene vorfand, in gesellschaftlichen Bereichen anzusiedeln, die durch die Wissenschaft noch nicht vereinnahmt waren. Die Quasivorgegebenheit des wissenschaftlichen Gegenstandes zu einem bestimmten historischen Zeitpunkt war immer Ausdruck und Produkt einer konkret tätigen Auseinandersetzung mit der Natur, in die die akkumulierte Erfahrung früherer Auseinandersetzungs-

prozesse einght, und die Form der Auseinandersetzung vollzog sich immer im Rahmen der jeweils arbeitsteilig organisierten Produktionsweise. Darin kann man die Voraussetzung für die Erklärung sehen, warum die Naturwissenschaften nicht einfach irgendwo beliebig ansetzen, sondern in ihrem Vorgehen durch den Gegenstandsbezug geprägt sind. Methodik und Methodologie als Teilaspekt der reflektierten Verallgemeinerung der Auseinandersetzungsformen mit der Natur müssen dann trotz des Widerstandes durch paradigmatische Stabilisierungstendenzen eine permanente Veränderung bzw. Differenzierung erfahren, weil sich die Auseinandersetzungsformen selbst ändern. Das bewußtlose Hinnehmen der Quasivorgegebenheit der Gegenstandsbereiche dürfte einer der Gründe sein, weshalb im Immanenzpositivismus das Erkenntnisprinzip formuliert werden konnte, daß alle Erkenntnis von der Analyse des G e g e b e n e n auszugehen hätte.

Diese Art der Gegebenheit möchte ich hier als Quasigegebenheit bezeichnen, weil durch den Verzicht auf die Rekonstruktion der empirisch operationellen Konstituiertheit von Gegenständen und durch den Verzicht auf die Reflexion der gesellschaftlichen Möglichkeit solcher empirischer Operationen die Gegenstände als schlechthin gegeben erscheinen können. Sie trifft bei ihrer a n a l o g i s i e r e n d e n Übertragung auf diagnostische Fragestellungen, also auf die Frage der Beurteilung von Dispositionen, auf ernsthafte Schwierigkeiten. Um diese konkretisieren zu können, bedarf es vorher einer allgemeinen, wissenschaftstheoretischen Strukturierung des Problems.

Wenn wir uns noch einmal erinnern, so wurde vorher gesagt, daß mit Dispositionsbegriffen die Potentialität von Gegenständen zum Ausdruck gebracht werden soll, sich unter angebbaren Bedingungen in angebbare Zustände zu transformieren. Die Transformationsbedingungen der Gegenstände sowie ihre verschiedenen Zustände und Bewegungsformen sollen in der Beobachtungssprache formuliert sein, von der angenommen wird, daß sie intersubjektiv vollständig verstehbar sei. Am Beispiel des in der Wissenschaftssprache einzuführenden Dispositionsbegriffes "Wasserlöslichkeit"

könnte man das Problem folgendermaßen ausdrücken: Es gibt einen oder mehrere Gegenstände im Zustand der Festigkeit, denen die Eigenschaft "Zucker" zukommt; die Gegenstände werden in Wasser gegeben und verändern dort ihren Zustand, indem sie sich auflösen. Kann man voraussetzen, daß die Begriffe "Zucker", "Wasser" und "Lösung" hinreichend genau festgelegt sind, so läßt sich ein Satz in der Beobachtungssprache formulieren:

"Wenn man dieses Stück Zucker in das Wasser wirft, so löst es sich auf". Gilt dieser Satz auch für andere Zuckerstücke, dann gelangen wir zu der Vermutung, daß "Wasserlöslichkeit" eine Eigenschaft von Zucker ist, die, weil sie nicht als solche beobachtbar ist, dem Zucker als Disposition zukommt.

Wie die Vermutung in eine allgemeingültige Aussage überführt werden kann, ohne daß die Operation des Hineinwerfens in das Wasser jeweils auszuführen ist, soll in diesem Zusammenhang nicht weiter interessieren. Wesentlich ist, daß sich die Bedeutung des Dispositionsbegriffes "Wasserlöslichkeit" nicht in den empirischen Operationen erschöpft bzw. in deren Abbildung in der Beobachtungssprache, sondern *d a r ü b e r h i n a u s*, unabhängig von einer künftigen Ausführung solcher Operationen, die allgemeingültige Invarianz der Transformation des Zuckers vom Festigkeitszustand in den Zustand des Aufgelöstseins im Wasser zum Ausdruck bringen soll.

Wenn man Dispositionen als invariante, potentielle Veränderungsform eines Gegenstandes begreift und mit dem Dispositionsbegriff die begriffliche Rekonstruktion einer bestimmten Transformationsinvarianz des Gegenstandes meint, läßt sich der Sinn des Charakteristikums der Nichtbeobachtbarkeit von Dispositionen aufklären. Gleichzeitig legitimiert sich damit nachträglich die von STEGMÜLLER angemahnte Notwendigkeit der Unterscheidung zwischen dem Fehlen beobachtungsmäßiger Kriterien für das Vorliegen eines Begriffes und der Undefinierbarkeit des Begriffes in der Beobachtungssprache. Denn unter der Voraussetzung, daß man Dispositionsbegriffe als begriffliche Rekonstruktionen einer bestimmten Transformationsinvarianz des Gegenstandes auffaßt, ist die Bedeutung

griffen durch zwei Kategorien von Operationen bestimmt:

1. Durch empirische Operationen, und zwar einerseits solche, die sich unmittelbar am Gegenstand vollziehen und andererseits jene, die der tätige Wissenschaftler durchführt, damit sich erstere realisieren; sie werden durch die Beobachtungssprache ausgedrückt: "Jemand wirft ein bestimmtes Stück Zucker in das Wasser und dieses löst sich darin auf". Die empirischen Operationen, die sich am Gegenstand unmittelbar vollziehen, stellen einerseits die objektive Grundlage für die Möglichkeit empirischer Erkenntnis dar, sind aber auf der Abbildungsebene der Beobachtungssprache zugleich Abstraktionen gegenüber der Gesamtheit des empirischen Geschehens - man könnte mit dem gleichen Vorgang ebenso gut die Disposition "Schweben" operationalisieren. Bereits auf dieser Ebene erweist sich, daß die Begriffe und Sätze der Beobachtungssprache nicht schlechthin auf Gegebenes deuten, sondern auf empirische Zustände und Veränderungsformen von Gegenständen, die durch die Beobachtungssprache begrifflich verallgemeinert werden können, weil sie auf der Ebene der empirischen Operationen verallgemeinerbar, d.h. reproduzierbar sind.

Die Frage nach erschöpfenden, beobachtungsmäßigen Kriterien für das Vorliegen eines Dispositionsbegriffes erübrigt sich dadurch, weil der Begriff z.B. "Wasserlöslichkeit" mitkonstituiert ist durch den verallgemeinernden und selektierenden Verwendungsaspekt des Beobachters. In die Begriffe der Beobachtungssprache gehen also zwei verschiedene empirische Momente ein, jene Operationen, die wir an den Gegenständen vollziehen, und jene, die selbständig in gesetzmäßiger Form ablaufen. Daneben sind aber bereits sprachkonstitutive Momente wirksam, die das Allgemeine an diesen Operationen, nämlich ihre Reproduzierbarkeit unter einem bestimmten Verwendungsgesichtspunkt zum Ausdruck bringen. Letzteres ist weder theoretisch ableitbar, noch aus der unmittelbaren Anschauung des erlebnismäßig Aufweisbaren zu gewinnen. Es repräsentiert vielmehr die intentionale Gerichtetheit der Auseinandersetzung mit der Natur, sei sie nun subjektiv motiviert oder gesellschaftlich bedingt, und

dürfte sich einer ausschließlich wissenschaftstheoretischen Formalisierung entziehen. Bei der Betrachtung der Beurteilung von Dispositionen in der Diagnostik wird dieser Problemkomplex eine außerordentlich wichtige Rolle spielen.

2. Durch nichtempirische Operationen, die wir in einer theoretischen Sprache formulieren. Durch sie wird der Dispositionsbegriff eigentlich erst in die Wissenschaftssprache als solcher eingeführt und das Problem der Nichtbeobachtbarkeit erst präzise bestimmbar. Hier geht es nicht mehr um die Frage, ob ein empirisches Phänomen - der Transformationsprozeß des Gegenstandes - durch die Beobachtungssprache zu verallgemeinern ist, sondern ob dem Transformationsprozeß unabhängig von seinem empirischen Auftreten Invarianz zukommt, bzw. ob die Beobachtungsaussage allgemeingültig ist. Die Nichtbeobachtbarkeit bezieht sich damit auf die *I n v a r i a n z* des Transformationsprozesses.

Da durch die Invarianzannahme per definitionem die Ebene des Beobachtungsmäßigen verlassen wird, kann die Entscheidung, ob Transformationsinvarianz vorliegt oder nicht, nicht mehr durch Rekurs auf beobachtungsmäßige Kriterien gefällt, sondern nur durch allgemeine Regeln festgelegt werden, die die Verbindung zwischen der Beobachtungssprache und der Invarianzannahme, nämlich der Annahme einer Disposition, herstellen. Diese Regeln betreffen gleichsam den Algorithmus der Reduktion des Dispositionsbegriffes auf Aussagen und Begriffe der Beobachtungssprache.

Auch hier erübrigt sich die Frage nach der Beobachtbarkeit der Disposition, weil die erwähnten Regeln die Verallgemeinerbarkeit der Beobachtungsaussagen formalisieren sollen und zwar im Hinblick auf jenen Verwendungsaspekt, unter dem auch die einzelnen Beobachtungsaussagen die konkreten empirischen Operationen verallgemeinern, d.h. unter dem die konkreten empirischen Operationen als untereinander vergleichbar bzw. als *g l e i c h e* Reproduktionen erscheinen.

Zusammenfassung der bisher erfolgten, allgemeinen Auseinandersetzung mit dem Dispositionsbegriff:

Mit der Einführung von Dispositionsbegriffen in die Wissenschaftssprache ist die Aufgabe zu lösen, reproduzierbare Transformationsprozesse von Gegenständen auf der Ebene der Beobachtungssprache zu verallgemeinern. Da dabei der Reproduktionsgesichtspunkt in Bezug auf einen bestimmten Verwendungszusammenhang die Grundlage der begrifflichen Verallgemeinerung liefert, und die dabei ablaufenden empirischen Operationen nicht als solche schlechthin stattfinden, sondern durch das Forschungsinteresse mitkonstituiert sind, stellt die Verallgemeinerung durch die Beobachtungssprache bereits einen Rekonstruktionsprozeß dar. Soweit die in der Beobachtungssprache verwendeten Begriffe empirisch bedeutungsvoll sind, ist diese empirische Bedeutung bereits mitvermittelt durch den selektierenden Betrachtungsaspekt, den man auf den Transformationsprozeß überträgt. Das heißt aber nicht, daß, eben weil sie auf die empirischen Operationen Bezug nimmt, die Beobachtungssprache eine bloß konstruktivistische Sprache sein kann.

Die eigentlich wichtigere Aufgabe ist die, die prinzipielle Reproduzierbarkeit der Transformationsprozesse in der Wissenschaftssprache begrifflich zum Ausdruck zu bringen und damit die Transformationsinvarianz als allgemeingültig festzulegen. Erst dann kann die Disposition als gegenstandskonstitutiv betrachtet werden. Dabei vollzieht sich ein erneuter Rekonstruktionsprozeß, durch den die beobachtungssprachlich repräsentierte Verallgemeinerung des Transformationsprozesses als unabhängig von zu benennenden Bedingungen anerkannt wird. Im Hinblick auf die Tatsache, daß das Ergebnis des Rekonstruktionsprozesses, der Dispositionsbegriff, trotz aller sprachkonstitutiven Vermitteltheit zu den empirischen Operationen durch die Operationen empirische Bedeutung erhält, bezeichne ich ihn in Anlehnung an SCHMIDT (1974) als R e k o n s t r u k t .

2.2 Orientierungsrahmen für die Rekonstruktion dispositioneller Schülereigenschaften: Allgemeine Bestimmung des Anwendungsbereiches pädagogischer Diagnostik

Mit der Einführung des Begriffes "Rekonstrukt" setzte ich den Ausgangspunkt für die Betrachtung des diagnostischen Prozesses in der Schule. Seiner Struktur nach umfaßt er die sprachliche Rekonstruktion von empirischen Operationen, die an den Gegenständen ausgeführt werden bzw. sich an den Gegenständen vollziehen. Orientiert an der empiristischen Wissenschaftstheorie, durchläuft der sprachliche Rekonstruktionsprozeß zwei kategorial unterschiedliche Ebenen, wenn das Rekonstrukt die Eigenschaften eines Dispositionsprädikates hat:

- a) Auf der Ebene der Beobachtungssprache wird das konkrete, empirische Geschehen durch Aussagen verallgemeinert und dadurch der intersubjektiven Vermittlung zugänglich gemacht.
- b) Auf der Ebene der theoretischen Sprache wird die Annahme begründet, daß das empirische Geschehen unter bestimmten Bedingungen grundsätzlich reproduzierbar ist bzw. seine sprachliche Rekonstruktion als allgemeingültig angesehen werden kann. Ergebnis dieses zweiten Rekonstruktionsprozesses ist ein Dispositionsprädikat, das die Invarianz der Gegenstandstransformation sprachlich repräsentiert.

Für die analogisierende Übertragung dieses Vorgehens auf die diagnostische Rekonstruktion dispositioneller Eigenschaften von Schülern sind nun zwei Probleme zu untersuchen:

1. Lassen sich Aussagen über Schülerverhaltensweisen in der gleichen Art als Beobachtungssprachliche Verallgemeinerungen von empirischen Operationen begreifen wie die Aussagen der physikalischen Beobachtungssprache ?
2. Können die Dispositionsprädikate als diagnostische Rekonstruktion von Schülereigenschaften die gleiche Struktur und Konstitution haben wie die physikalischen Dispositionsprädikate ?

Die Untersuchung der Übertragbarkeit wird mit der Beantwortung

der zweiten Frage beginnen, weil zum einen die Klärung der Schwierigkeiten, die sich aus der Bestimmung einer diagnostischen Beobachtungssprache ergeben, die Analyse der Funktion diagnostischen Dispositionsprädiakte voraussetzt. Zum zweiten entspricht dieses Vorgehen der praktizierten Festlegung von Qualifikationsmustern im Ausbildungsbereich, wo erst nach der Festlegung von zu erreichenden Dispositionen im Fortgang der Ausbildung der empirische Gehalt von Qualifikationsmustern zu finden versucht wird.

Sieht man von den verschiedenen Verwendungskontexten diagnostischer Aussagen ab, so kann man einen Minimalkonsensus in Bezug auf die gnoseologischen Vorbedingungen diagnostischer Forschung in der Weise herstellen, wie SCHMIDT (1974) dies an vier Bestimmungen durchführt:

1. Die empirische Diagnostik bemüht sich um das Erkennen von dispositionellen psychischen Eigenschaften als Bindeglieder zwischen Umgebungseinwirkungen und dem Verhalten.
2. Über die Natur dieser Dispositionen ist a priori zu sagen, daß sie unabhängig vom Bewußtsein des Forschers oder Diagnostikers als objektiv-reale Tatbestände, gebunden an materielle Prozesse und Strukturen im Zentralnervensystem, existieren.
3. Da die Eigenschaften von Objekten, dialektisch betrachtet, stets Produkte von Wechselwirkungen sind, ist auch die Rekonstruktion psychischer Eigenschaften eine Funktion des mit der Forschung oder Diagnostizierung realisierten Interaktionsvorganges.
4. Der empirische Forschungs- oder Diagnostizierungsprozeß rekonstruiert die hier vorgefundene Realität in theoretischen Sprachen, in denen sich relative Wahrheiten über die Natur der dispositionellen Eigenschaften abbilden (ebd., S. 22).

Mit diesen vier Bestimmungen ist allerdings noch keine hinreichende Spezifikation der diagnostischen Fragestellung erreicht; solange sie weder in ihrer gegenseitigen Abhängigkeit betrachtet werden noch die Bestimmungen 1. und 3. näher untersucht sind, gelten sie in gleicher Weise für die physikalischen Dispositionsbegriffe.

Der entscheidende Problemkomplex verbirgt sich hinter dem in der 3. Bestimmung gebrauchten Begriff "Objekt"; dem Objekt sollen

diagnostisch rekonstruierte Eigenschaften als Produkte von Wechselwirkungen zugeordnet werden, wobei die Wechselwirkung zwischen dem "Objekt" - hier dem Schüler - und dem Forscher bzw. Diagnostiker stattfinden.

Unbezweifelbar ist der Diagnostiker an der Rekonstruktion von Dispositionen von Schülern interessiert und damit an der Findung von Eigenschaften des "Objektes". Andererseits beginnt aber die diagnostische Rekonstruktion dispositioneller Eigenschaften mit der Auslösung von Verhaltensweisen durch festgelegte, situative diagnostische Bedingungen. Dadurch sind die zu rekonstruierenden Dispositionen über den Gesamtkomplex der diagnostischen Bedingungen und der Auseinandersetzung damit vermittelt. Daraus ergibt sich die Konsequenz, daß der materielle Untersuchungsgegenstand des Diagnostikers nicht das Objekt "Schüler" ist, sondern das vergegenständlichte Resultat der Auseinandersetzung des Schülers mit den diagnostischen Bedingungen. Der Schüler fungiert nur insofern als Objekt mit dispositionellen Eigenschaften, als er der Produzent der Resultate ist. Die verschiedenen empirischen Operationen, die er im Vollzug der Auseinandersetzung mit den diagnostischen Bedingungen durchführt, sind zwar irgendwie im Auseinandersetzungsergebnis enthalten, treten dort aber nicht mehr manifest in Erscheinung und können deshalb aus diesen allein nicht mehr rekonstruiert werden.

Realiter konstituieren aber die verschiedenen vom Schüler vollzogenen empirischen Operationen das Auseinandersetzungsergebnis. Ihre Vernachlässigung bei der diagnostischen Rekonstruktion von Dispositionen würde der Eigenschaftsbestimmung bei Schülern zum Zweck der Beurteilung ihrer Eignung für weiterführende Ausbildungsalternativen den Boden der empirischen Bedeutungshaftigkeit entziehen. Das hat seinen Grund darin, daß der Prozeß der Rekonstruktion von dispositionellen Eigenschaften "außerdem einem anderen Interaktionssystem (angehört; d. Verf.) als das untersuchte Objekt (Schüler; d. Verf.), dessen dispositionelle Eigenschaften zu erschließen sind" (SCHMIDT, 1974, S. 23). SCHMIDT schematisiert die Rekonstruktion dispositioneller Eigenschaften

im Bezugsrahmens des Zusammenhanges zweier Interaktionssysteme, das des untersuchten Individuums und das des Forschungsprozesses:

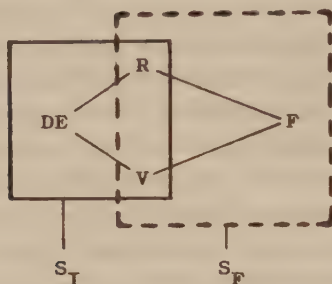


Abb. 1: Schematisierung der Rekonstruktion dispositioneller Eigenschaften nach SCHMIDT (1974, S. 23).

Gemäß dieses Bezugsrahmens ist die Rekonstruktion von Dispositionen abhängig von der Verarbeitung der Reizlage (R) durch das untersuchte Individuum, sowie - und dies wird von SCHMIDT nicht eigens thematisiert - von der weiteren Verarbeitung der Verhaltensresultate (V), die im Verlauf der Bearbeitung von Prüfungsaufgaben regelnd zurückwirken auf den Lösungsversuch solcher Prüfungsaufgaben oder aber auf den Lösungsversuch anderer Aufgaben. Dieser Operationenkomplex ist im Interaktionssystem des untersuchten Individuums (S_I) lokalisiert und wird von der hypostasierten dispositionellen Eigenschaft (DE) gesteuert. Zum anderen ist die Rekonstruktion abhängig vom Interaktionssystem des Forschers (S_F), in dem der Forscher (F) den Datenkomplex selektiv reduziert, Klassifikationen usw. vornimmt, und durch die Kontingentierung von Klassen von R und V zu Bestimmungen von dispositionellen Eigenschaften zu gelangen versucht. Die Rekonstruktion dispositioneller Eigenschaften ist damit ebenso abhängig von der Spezifität der Verarbeitung der Reizlage und der Verhaltensresultate durch den Diagnostiker.

Daß beide Arten der Verarbeitung von Reizlage und Verhaltensresultat, die des untersuchten Individuums und die des Diagnos-

tikers bzw. Forschers nicht nur verschieden sein können, sondern grundsätzlich verschieden sein müssen, bedürfte an sich keiner besonderen Erwähnung. Es ergibt sich aber daraus die Konsequenz, daß man in der diagnostischen Rekonstruktion der hypostasierten dispositionellen Eigenschaften von Schülern zu qualitativ unterschiedlichen Ergebnissen gelangt, je nach dem, welche Art der Verarbeitung man zur Grundlage der Rekonstruktion macht.

2.2.1 Zwei Beispiele für die Rekonstruktion dispositioneller Eigenschaften

Erstes Beispiel

Die detaillierte Klärung der Unterschiede in den Ergebnissen steht wiederum mit der inhaltlichen Ebene in Zusammenhang, auf der die Verarbeitung der Reizlage und der Verhaltensresultate untersucht wird. Mit Verarbeitung der Reizlage bzw. des Verhaltensresultates, die in ihrer Beziehung zueinander die Grundlage für die diagnostische Rekonstruktion von dispositionellen Eigenschaften darstellen, ist allgemein ihre Kodierung durch den Organismus gemeint. In Bezug darauf führt bereits die unterschiedliche, physiologische Wahrnehmungsstruktur des untersuchten Individuums bzw. die des Diagnostikers oder Forschers oder aber der physiologisch nachweisbare Einfluß von organisch umgesetzten Erwartungen und Motiven in der Testsituation für das untersuchte Individuum bzw. in der Testkonstruktionssituation für den Diagnostiker zu jeweils verschiedenen Rezeptionen der Prüfungsaufgabe und der Lösungen dieser Prüfungsaufgabe.

Möglicherweise kann die Kenntnis des Einflusses von physiologischen Faktoren im Diagnostiker auf seine Klassifikation und Interpretation von Testitems und der entsprechenden Lösungsversuche vernachlässigt werden. Mit Sicherheit wäre sie erst dann zu berücksichtigen, wenn die Wirksamkeit von anderen Faktoren

wie gesellschaftlich vermittelte Leistungsstereotype, berufssozialisationsbedingte Erwartungen an zu überprüfende Qualifikationen von Schülern, privatisierte Normvorstellungen in Bezug auf Ausprägungsgrade schulischer Leistung usw. geklärt ist. Die Reflexion ihrer Wirksamkeit im Interaktionssystem S_I kann aber durchaus Konsequenzen für die diagnostische Rekonstruktion von dispositionellen Eigenschaften von Schülern nach sich ziehen.

So zeigen z.B. die Untersuchungen zum Verhältnis von "vigilance" und "attention", deren Bedeutung für das Verhalten in Prüfungssituationen bis heute nicht ansatzweise erforscht ist, daß Rückmeldungen über Teilerfolge bei der Lösung von Aufgaben den Wachsamkeitsabfall in der Zeit verringern (MACKWORTH, 1970). Ein Diagnostiker, der das Phänomen kennt, könnte Klassen von Testaufgaben mit Rückmeldefunktion bevorzugen, um den Wachsamkeitsabfall auf ein Minimum zu reduzieren und damit die zu rekonstruierende Disposition (z.B. "konvergentes Denken") präziser eingrenzen. Er käme damit dem konkreten, empirischen Geschehen, das er hinter der Bezeichnung "konvergentes Denken" vermutet und soweit es überhaupt durch diagnostische Anregungsbedingungen repräsentiert bzw. induziert werden kann, in der Rekonstruktion näher als jener Diagnostiker, der in Unkenntnis der Kompetenz des Sachverhaltes dem Schüler zur Überprüfung des konvergenten Denkens eine Reihe von unverbundenen Testaufgaben mit Richtig-Falsch-Antworten vorsetzt.

Letzterer müßte das Ergebnis der Prüfung als Ausprägungsgrad der genannten Eigenschaft interpretieren und somit als Disposition, ohne die implizit wirksame und vielleicht dispositionelle Ermüdungstendenz des Schülers in Rechnung stellen zu können. So könnten z.B. Unterschiede zwischen Schülern in Bezug auf einen Test zur Untersuchung des konvergenten Denkens zum Großteil oder allein auf dispositionelle Unterschiede im Wachsamkeitsabfall bei der Bearbeitung von Testitems zurückzuführen sein, die keine Rückmeldung von Teilerfolgen zulassen.

An der zuletzt genannten Gegenüberstellung beider Typen von Diagnostikern kommt zum Ausdruck, daß sie trotz der Verwendung der gleichen Bezeichnung für eine Disposition unterschiedliche Interpretationen mit dem Rekonstrukt "konvergentes Denken" verbinden, weil sie sich in unterschiedlicher Weise auf das Interaktionssystem S_I beziehen. Letzterer reduziert das Interaktionssystem S_I auf den Faktor DE und ist daher gezwungen, eine dispositionelle Eigenschaft des Schülers in ihrer Ausprägung aus der Menge der wiederholt aufgetretenen Kontingenzen zwischen bestimmten R und V zu "erschließen". Als Bestimmungsmerkmale für die Disposition dienen ihm ausschließlich die von ihm gewählten Klassifikationen von Testitems; für die verschiedenen R könnte die Klassifikation "Items, die konvergentes Denken provozieren" heißen und für die verschiedenen Reaktionen darauf "Richtig" bzw. "Falsch".

Damit wird aber das Interaktionssystem S_I aus der Untersuchung ausgeklammert. Die Tatsache, daß der Schüler die diagnostische Reizlage nicht mehr im Sinne des Testkonstrukturs kodifiziert, erscheint nur noch in der falschen Beantwortung von Prüfungsaufgaben und somit als Verringerung des Ausprägungsgrades der dispositionellen Eigenschaft. Abgesehen davon, daß der Schüler unter veränderten Bedingungen, z.B. bei verringerten Wachsamkeitsabfall nach einer dazwischengeschobenen Pause während des Testens, zur Lösung der Prüfungsaufgaben hätte gelangen können, kann die Verallgemeinerung des gelieferten Lösungsergebnisses durch die Dimensionierung "Richtig-Falsch" nicht mehr zu den empirischen Prozessen in Beziehung gesetzt werden, die bei der Bearbeitung der Prüfungsaufgabe hätten untersucht werden müssen. Denn was als Richtiglösung einer Prüfungsaufgabe anzusehen ist, hängt unter den gegebenen Bedingungen in keiner Weise von den verschiedenen psychischen Prozessen ab, die zur Richtiglösung führen. Das gleich gilt auch für die falsche Beantwortung einer Aufgabe.

Genau genommen kann daher die dispositionelle Eigenschaft "Konvergentes Denken" nur noch als Reaktionstendenz bestimmt werden,

auf die vom Diagnostiker vorgegebene und festgelegte Klasse von Testitems in der vom Diagnostiker als richtig angesehenen Weise zu antworten. Aussagen über die empirische Konstitution der so "erschlossenen" Dispositionen sind auf der Basis dieser Feststellungen nicht möglich, obwohl die Klassifikationen der Testitems und der verschiedenen Reaktionen darauf nicht nur Konventionen darstellen. Sie können auch eine gewisse intuitive Erfahrungsplausibilität beanspruchen. Diese erschöpft sich allerdings nur in dem Wissen, daß die ausgewählte Klasse von Testitems Antworten durch die Schüler in der Weise zu provozieren vermag, daß mit ihrer Hilfe zwischen den Schülern unterschieden werden kann.

Im Gegensatz dazu bezieht der zuerst erwähnte Diagnostiker die Reizlage und Verhaltensresultate in die Untersuchung des Interaktionssystems S_I mit ein, indem er die Kodifizierung der Reizlage durch das untersuchte Individuum als Größe des Einflusses auf den Lösungsversuch problematisiert. Er versucht damit die Tatsache in die Bestimmung der dispositionellen Eigenschaften mitaufzunehmen und zu strukturieren, daß das untersuchte Individuum im Vollzug des Lösungsprozesses Operationen ausführt, die jenen des Diagnostikers formal gleichzusetzen sind, nämlich solche der Klassifikation und Interpretation der Testitems und seiner eigenen Reaktionen darauf.

Zwar ist auch dieser Ansatz von der Gefahr der naiv-empiristischen Verkürzung nicht befreit, wenn - um am Beispiel zu bleiben - die Art der an das untersuchte Individuum zurückgemeldeten Information über Teilerfolge in der Bearbeitung der Prüfungsaufgaben ausschließlich vom Diagnostiker bzw. Forscher im Hinblick auf die erfolgreiche Zuendeführung des Lösungsprozesses strukturiert wird. Nachwievor erfährt die Auseinandersetzung des untersuchten Individuums mit den diagnostischen Anregungsbedingungen ihre Subsumption unter die im Interaktionssystem S_F festgelegten Klassifikations- und Interpretationsgesichtspunkte. Deren empirische Bedeutung im Sinne der empirischen Operationen, die der Schüler bei der Auseinandersetzung ausführt, ist noch

zu erforschen. Es wird aber dadurch - sofern man geneigt ist, das Vorläufige an diesem Verfahren ernst zu nehmen - der diagnostischen Forschung die Möglichkeit eröffnet, die Urteilsbildung als Rekonstruktion der empirischen Operationen zu begreifen, die als komplexes Zusammenwirken im Interaktionssystem S_I den Vollzug der Bearbeitung der Prüfungsaufgaben darstellen. Nur aus diesem Grund stimmt SCHMIDT der Bildung von Rekonstruktionen mit Hilfe von Operationaldefinitionen als einem ersten, vorläufigen aber durchaus zweckmäßigen und legitimen Weg zu, der im Fall der empirischen Überprüfung theoretischer Begriffsbildung sogar unerlässlich ist (ebd., S. 24).

Ein weiterer Unterschied gegenüber dem vorher besprochenen Vorgehen besteht darin, daß die kontingente Beziehung zwischen der Reizlage und dem Verhaltensresultat über die zurückgemeldeten Informationen vermittelt ist und diese Vermittlung in der diagnostischen Rekonstruktion der dispositionellen Eigenschaft explizit ausgewiesen sein muß. Durch die explizite Berücksichtigung des dispositionskonstitutiven Faktors der informativen Rückmeldung kann die extrapolierte Eigenschaft "Konvergentes Denken" nicht mehr ausschließlich als privates Charakteristikum des untersuchten Individuums begriffen werden. Sie ist die sprachliche Verallgemeinerung der festgestellten, invarianten Beziehung zwischen der Klasse der Reizlagen und der Klasse der Verhaltensresultate, wenn vom Interaktionssystem S_F bestimmte, operative Einfriffe - hier: Rückmeldungen über Teilerfolge in der Bearbeitung von Prüfungsaufgaben - in den Lösungsprozeß erfolgen. Das hat wiederum zur Konsequenz, daß die Struktur des Rekonstrukts "Konvergentes Denken" nicht mehr als bloß quantifiziertes Abstraktum in Erscheinung treten kann, sondern nur im Kontext der Operationen Bedeutung hat, die im Interaktionssystem S_F vorgenommen werden.

Dieses Vorgehen entspricht der alltäglichen Beurteilung von Schülern weit mehr: Soll z.B. ein Schüler eine mündliche Frage des Lehrers beantworten und gerät er bei der Problemlösung in Schwierigkeiten, so kann bereits ein Denkanstoß des Lehrers zur

Behebung der Schwierigkeiten führen. Die Hilfestellung des Lehrers ist konstitutiv für die Bewältigung der Aufgabe durch den Schüler und muß infolgedessen auf der diagnostischen Rekonstruktionsebene in die Beurteilung von Eigenschaften des Schülers eingehen. Eine abstrakt quantifizierende Beurteilung der erfolgreichen Aufgabenbewältigung als Ursache zugrundegelegten dispositionellen Eigenschaft ist in diesem Fall nicht nur deshalb nicht möglich, weil die konstitutive Hilfestellung nicht verallgemeinerbar ist, d.h. weil sie aus der Spezifität des interaktiven Geschehens heraus entwickelt wurde bzw. auf den spezifischen Schüler zugeschnitten ist. Sie ist v.a. deshalb unangebracht, weil damit die Effektivität der Hilfestellung implizit und unreflektiert im quantitativen Maßstab als mitbeurteilte enthalten wäre, d.h., das interaktive Geschehen auf einer Dimension abgebildet würde. Daß gegen dieses Prinzip insbesondere dann verstoßen wird, wenn die als Hilfestellung gemeinte Intervention des Lehrers keine erfolgreiche Bewältigung der Aufgaben hervorruft und der konstatierte Mißerfolg auf das Konto fehlender Fähigkeiten des Schülers verbucht wird, ist ein Problem, das auf der vorliegenden Ebene der Betrachtung allein nicht zu analysieren ist.

Am Beispiel des möglichen Einflusses von physiologisch kontrollierbaren Faktoren wie "Wachsamkeit" auf die Bearbeitung von Prüfungsaufgaben, mit denen "konvergentes Denken" untersucht werden soll, war zu zeigen, daß die diagnostische Rekonstruktion von dispositionellen Eigenschaften des untersuchten Individuums in qualitativ unterschiedlichen Ergebnissen mündet, je nachdem, ob dieser Teilaspekt des interaktiven Geschehens im Interaktionssystem S_I durch den Diagnostiker oder Forscher reflektiert wird oder nicht. Beide Typen von Diagnostikern lassen sich daher insofern unterscheiden, als sie bestimmte Bedingungen der Verarbeitung der Reizlage und des Verhaltensresultates durch das untersuchte Individuum zur Kenntnis nehmen und als konstitutiv für die Rekonstruktionsbildung betrachten oder aber die Rekonstruktion der dispositionellen Eigenschaft nur auf der

Grundlage ihrer jeweils S_I -unbezogenen Strukturierung von Reizlage und Verhaltensresultat vollziehen.

Zweites Beispiel

Wie bereits erwähnt, kann aber auch die explizite Berücksichtigung von S_I in einem naiv-empiristischen Rekonstruktionsversuch dispositioneller Eigenschaften münden, wenn die Aussagen über den Verlauf und die Bedingungen des Aufgabenlösungsverhaltens aus einer oberflächlichen Beschreibung des Interaktionsgeschehens in S_I gewonnen, als Hypothesen ausgegeben und formalmethodisch als allgemeingültig zu bestätigen versucht werden. Kennzeichnend für solche Vorgehensweisen ist die analytisch-begriffliche Isolierung von bestimmten Wirkgrößen des interaktiven Geschehens mit Hilfe von Kategorien, die auf der Ebene der oberflächlichen Betrachtung anschauliche Bedeutung zu haben scheinen, und der Versuch ihrer Synthese durch formalmethodische Bezugssysteme, die von Anfang an zur Auswahl stehen und sich zunächst nur durch ihre konventionalisierte Verwendung in der Wissenschaft legitimieren.

Hierunter fallen alle Formalisierungsversuche, die auf der statistischen Korrelation zwischen Variablen aufbauen und die experimentellen Versuche, den Effekt von Wirkungsgrößen durch deren systematische Variation und gleichzeitiger Konstanthaltung und Kontrolle von sog. Störeffekten aufzuzeigen. In der Verfolgung des Zieles, zunehmend präzisere Beschreibungen und Erklärungen von Verhaltenstrukturen zu liefern, gerät man dabei in das Dilemma, die sukzessive Einschränkung des Anwendungsbereiches der Erklärungsansätze durch Erweiterungen des Variablenkataloges oder durch ad-hoc-Anpassungen der Erklärungsansätze mittels Zusatzannahmen ausgleichen zu müssen.

Bei solchen Untersuchungen des Interaktionssystems wird ein naiv-empiristisches Vorgehen praktiziert, d.h., die Untersuchungen beschränken sich auf die Kontingentierung von Reizlageklassen und Verhaltensresultatklassen durch methodische Reduktionsmodelle. Die vollzogene Klassifizierung kann jeweils nicht

durch die Analyse des Prozesses der Verarbeitung der diagnostischen Situation durch das untersuchte Individuum inhaltlich mitbegründet werden. Aus diesem Grund fungiert das methodische Reduktionsmodell selbst als Kriterium für die Begrenzung des Variablenkataloges bzw. der Zusatzannahmen. So muß man z.B. bei der Anwendung der Faktorenanalyse auf einen Test die Voraussetzung machen, daß die einzelnen Items eine hohe Varianz besitzen und damit im Idealfall mittlere Lösungswahrscheinlichkeit zeigen. Aufgaben mit extremer Lösungswahrscheinlichkeit müssen entsprechend dieser Forderung aus dem Test beseitigt werden, da sie nicht die Bedingungen des methodischen Reduktionsmodells "Faktorenanalyse" erfüllen. Es sei nur nebenbei erwähnt, daß diese Forderung im Widerspruch zu bestimmten pädagogischen Intentionen stehen kann. Zur Erhöhung der Lernmotivation und zur Verbesserung der Unterrichtseffektivität durch die Ermöglichung einer häufigeren Erfolgsrückmeldung ist es z.B. sinnvoll, Items mit hoher Lösungswahrscheinlichkeit darzubieten.

In Bezug auf die diagnostische Rekonstruktion von dispositionellen Eigenschaften führt die Anpassung der Klassifikation von Reizlage und Verhaltenstestresultat an die immanenten Bedingungen des methodischen Modells zu unauflösbaren Schwierigkeiten, weil die Verarbeitung der diagnostischen Situation durch das untersuchte Individuum nur aus der Perspektive des Interaktionssystem S_F analysiert wird. Zur Veranschaulichung greife ich auf die Untersuchungen von WALLACH & KOGAN (1965) über den Zusammenhang von Testintelligenz und Testkreativität unter verschiedenen situativen Bedingungen zurück. Sie konnten nachweisen, daß sich die Interkorrelationen innerhalb der verwandten, unterschiedlichen Intelligenztests verringern, wenn die Testaufgaben in einer vergleichsweise streßfreien Situation dargeboten werden. Vernachlässigt man bei der Interpretation dieser Befunde die noch zu untersuchenden Prozesse im Interaktionssystem S_I vollkommen, so muß man, wenn aus den vorliegenden Ergebnissen mittels Faktorenanalyse auf zugrundeliegende dispositionelle Eigenschaften "geschlossen" werden soll, daraus zwei alternative, widersprüchliche Konsequenzen ziehen:

1. Wird die relativ streßfreie Darbietung der Testaufgaben aus irgendetwelchen Gründen (z.B. zur Überwindung der Testangst) zugelassen, dadurch die Lösungswahrscheinlichkeit erhöht und die Varianz verringert, so verändert sich dadurch die Faktorenstruktur gegenüber der restringierten Darbietung. Man wäre daher gezwungen, die Unterschiede zwischen den jeweiligen Faktorenstrukturen durch die Unterschiede zwischen den Darbietungssituationen zu interpretieren und die Darbietungssituation als konstitutives Moment für die Bildung der Rekonstrukte "Faktoren" anzuerkennen. Dies steht aber im Gegensatz zum Anspruch des Tests, "die" dispositionelle Eigenschaft - z.B. Intelligenz - r e l i a b e l abzubilden. Man würde deshalb den Test als unbrauchbar bezeichnen bzw. durch die Selektion von Aufgaben mit extrem hoher Lösungswahrscheinlichkeit zu "verbessern" suchen.

2. Man versucht durch die Entscheidung für die restringierte Darbietungssituation den ursprünglichen Interpretationsanspruch zu retten, zumal sich dadurch die Varianzen für die Testaufgaben erhöhen und den Modellbedingungen der Faktorenanalyse entgegengekommen wird. Man hätte sich aber damit gleichzeitig die Chance genommen, die Beziehung zwischen den verschiedenen Befunden durch jene Vorgänge zu klären, die ganz offensichtlich für die Divergenz zwischen den Ergebnissen mitverantwortlich zu machen sind, nämlich durch die Darbietung der Testaufgaben in der streßfreien Situation und der restringierten Situation.

Da außerdem die ursprüngliche Interpretation des Tests als Indikator von Intelligenz nur in der restringierten Darbietungssituation aufrechtzuerhalten ist, geht die Darbietungssituation implizit in die Definition von Intelligenz ein. Damit wäre aber der Gegensatz zwischen dem Anspruch einer situationsneutralen Erfassung d e r Intelligenz und der realiter situationsspezifisch erfaßten Intelligenz nur durch die Ignoranz gegenüber den anderen situativen Bedingungen der Intelligenzerfassung verdeckt.

Der Vollständigkeit halber ist der Erwähnung der Studie von WALLACH & KOGAN hinzuzufügen, daß die Autoren keinen faktorenanalytischen Rekonstruktionsversuch von dispositionellen Eigenschaften unternehmen. Sei wollen ja gerade in ihrer Auseinandersetzung mit den beiden Rekonstrukten "Intelligenz" und "Kreativität" aufzeigen, daß die situativen Bedingungen der Darbietung von Tests qualitativ unterschiedliche Interpretationen der Testergebnisse erforderlich machen. Die Studie im Rahmen der Betrachtung der diagnostischen Rekonstruktion von dispositionellen Eigenschaften anzuführen und ihre Ergebnisse unter die Lupe einer potentiellen Faktorenanalyse zu nehmen, ist aber deswegen sinnvoll, weil sich an ihnen die Interpretationsambivalenz von Testergebnissen besonders verdeutlicht und weil "die Faktorenanalyse eine der gängigsten Methoden zur Auffindung und Beschreibung von latenten Eigenschaftsdimensionen ist" (FISCHER, 1974, S. 77).

Sie ist allerdings nicht mehr als eine Methode, mit der eine Menge von Testaufgaben danach klassifiziert wird, ob diese in gleicher Weise die Personen einer untersuchten Population zu unterscheiden vermögen. Über die inhaltliche Bestimmung der Unterscheidungsdimensionen ist damit noch nichts ausgesagt. Sie kann daher, wenn man ihre Funktion innerhalb eines empiristischen Wissenschaftsansatzes sieht, lediglich die Aufgabe der Bestätigung einer vor ihrer Anwendung vorgenommenen Klassifikation der Testaufgaben haben. Ob die so bestätigte Klassifikation empirische Bedeutung hat oder nicht, ist von der Durchführung der Faktorenanalyse völlig unabhängig, da durch sie nur eine rechnerische Reduktion und Strukturierung vorhandener Daten erfolgt. Die empirische Bedeutung der Klassifikation muß schon vorher abgeklärt sein und kann sich durch die Anwendung des faktorenanalytischen Verfahrens nur in dem eingeschränkten Sinn als richtig erweisen, wie die klassifizierten Testaufgaben als Indikatoren einer dispositionellen Eigenschaft aufgefaßt werden können, die in den untersuchten Individuen als graduell abgestuft angenommen wird.

Damit hängt also die empirische Interpretationsmöglichkeit der erhaltenen Faktorenstruktur ausschließlich von der empirischen Interpretationsmöglichkeit der Klassifikation von Testaufgaben ab, die unabhängig von der Anwendung der Faktorenanalyse festliegen muß und durch diese mehr oder weniger ausreichend rechnerisch rekonstruiert werden kann. Voraussetzung dafür ist allerdings, daß die rekonstruierten Daten jene Skaleneigenschaften besitzen, die von den mathematischen Operationen in der Faktorenanalyse gefordert werden (vgl. KALVERAM, 1970).

Nun zeigen aber die Untersuchungen von WALLACH & KOGAN, daß die Bestimmung der empirischen Bedeutung von Testaufgaben eng zusammenhängt mit den Bedingungen, unter denen die Aufgaben bearbeitet werden. Dieses Problem wird in der Testtheorie unter dem Gesichtspunkt der Reliabilität von Tests thematisiert und wurde in der psychologischen Diagnostik intensiv untersucht. Bedauerlicherweise geht die testtheoretische Sichtweise aber davon aus, daß die Empfindlichkeit der Testinstrumente gegenüber situativen Einflüssen eine Eigenschaft des Testinstrumentes selber ist und nicht das Resultat der Auseinandersetzung des untersuchten Individuums mit der Testsituation. Aus der Tatsache des experimentell nachgewiesenen Einflusses von bestimmten situativen Momenten der gesamten Testdurchführung wie Zimmertemperatur, Helligkeit, Anordnung der Personen im Raum, Anzahl der Aufsichtspersonen im Raum, unkontrollierte Hilfestellung durch den Versuchsleiter usw. wurde der Schluß gezogen, die Testsituation insgesamt zu standardisieren, um die Wirksamkeit solcher Störeinflüsse zu vermeiden bzw. zu vereinheitlichen. Hinter dieser Maßnahme verbirgt sich ein naiv-funktionalistisches Modell menschlicher Verhaltenssteuerung mechanistischer Provenienz, wonach die verallgemeinerbaren Bedingungen menschlichen Verhaltens optimal zu instrumentalisieren sind und die dennoch auftretenden Unterschiede auf die Wirksamkeit von individuumspezifischen Dispositionen zurückgeführt werden müssen. Es scheint unter Zugrundelegung dieses funktionalistischen Modells nahezuliegen, das Phänomen, daß die relativ hohen Interkorrela-

tionen zwischen den Intelligenztests bei restringierter bzw. standardisierter Darbietung und die vergleichsweise niedrigen bei liberalisierter Darbietung durch die Wirksamkeit eines Komplexes von Störvariablen zu erklären, die im zweiten Fall zur situationskomponentenspezifischen Auseinandersetzung mit der Testsituation führt. Die Benennung der spezifischen Komponenten ist aber solange nicht möglich, wie die inhaltliche Dimensionierung der Standards in der restringierten Darbietung nicht angegeben werden kann. Letzteres ist wiederum deshalb nicht möglich, da die Standardisierung nur negatorisch zu begründen ist, nämlich nur in Bezug auf alle situativen Komponenten, die als Störkomponenten zwar beliebig nachweisbar aber in ihrer Wirkung nicht erklärbar sind. In der Regel wird daher die inhaltliche Dimensionierung der als verhaltensprovokativ betrachteten, standardisierten diagnostischen Situationen aus der Klassifikation der Verhaltensresultate bezogen und zwar nach dem Prinzip: ein bestimmtes Item ist deshalb ein Intelligenzitem, weil das provozierte Verhalten unter dem Gesichtspunkt der Intelligenz klassifizierbar ist. M.a.W.: Der Bestimmungsversuch der inhaltlichen Bedeutung der Testaufgaben läuft Gefahr, zu einem *circulus vitiosus* zu geraten, wenn das provozierte Verhalten nicht unabhängig vom provozierenden Item klassifiziert und bestimmt werden kann.

2.2.2 Formale Bestimmung des Anwendungsbereiches

Für die Aufklärung der Struktur von diagnostisch relevanten Dispositionsbegriffen sollen die zwei erwähnten Beispielsfälle unter einer verallgemeinerten Perspektive betrachtet werden. Dabei ist die nachwievor im Hintergrund stehende Frage zu beantworten, wie der Anwendungsbereich pädagogisch-psychologischer Diagnostik im Kontext der selektionsorientierten Form organisierten Lernens bestimmt werden kann. Da es in der selektionsbezogenen Diagnostik in der Schule in erster Linie darauf an-

kommen muß, in ihrem Vorgehen auf einer adäquaten, diagnostischen Rekonstruktion von dispositionellen Eigenschaften aufbauen zu können, liefert die Aufklärung der Struktur diagnostisch relevanter Dispositionsbegriffe die systematische Vorklärung des intendierten Anwendungsbereiches pädagogisch-psychologischer Diagnostik.

Die Erörterung der zwei Beispielsfälle ergab, daß der Versuch, die konstatierbare, formale Beziehung zwischen einer Klasse von Reizlagen und einer Klasse von Verhaltensresultaten empirisch zu interpretieren, in Widersprüche führt bzw. ein zirkuläres Vorgehen darstellt. Konnte im ersten Fall die empirische Bedeutung der Beziehung zwischen R und V dann nicht untersucht werden, wenn das interaktive Geschehen S_I außerhalb des Reflexionshorizontes des Forschers bzw. Diagnostikers blieb und die Klassifizierung und Kontingentierung von R und V ausschließlich im Kontext des Interaktionssystems S_F erfolgte, so vergrößerte der darüber hinausgehende Versuch, die Klassifizierung der Reizlage (Testaufgaben) durch Faktorenanalyse zu objektivieren, zusätzlich die Interpretationsschwierigkeiten. Die Klassifizierung mußte an die Modellbedingungen der Faktorenanalyse angepaßt werden. Ungeklärt bleibt auch dabei das Problem der Begründung der unabhängig von der Faktorenanalyse vorzunehmenden Klassifikation bzw. der Bestimmung ihrer empirischen Bedeutung. Letztlich wird ja mit der Faktorenanalyse nicht mehr erreicht als eine Formalisierung des Versuchs der Gruppierung von verschiedenen Testaufgaben, deren Gruppierungsgesichtspunkt einer nachträglichen Interpretation bedarf.

Gleiches gilt in Bezug auf das Interpretationsproblem auch für die Verfahren der Itemselektion nach der Lösungswahrscheinlichkeit oder nach der Trennschärfe, nach Validitätskoeffizienten oder nach dem Prinzip der Diskriminanzanalyse. Mit all diesen Verfahren wird im Endeffekt die Hypothese überprüft, ob eine vorläufige und intuitiv getroffene Auswahl von Testaufgaben die Personen einer Population hinsichtlich der unterstellten Dimensionierung der Testaufgaben maximal diskriminiert. Die Diskri-

minationsleistung der Testaufgaben läßt sich zwar dadurch ermitteln; die Festlegung der inhaltlichen Dimensionierung hängt aber nachwievor ausschließlich vom Forscher oder Diagnostiker bzw. von der Forschergemeinschaft oder allgemein von der Gemeinschaft ab, die ein Interesse an der Entwicklung und Anwendung von "objektiven" Beurteilungsverfahren hat.

Das Unternehmen, die beiden behandelten Beispiel als Typen für die Bestimmung des intendierten Anwendungsbereiches einer pädagogisch-psychologischen Diagnostik einzuführen, muß also erfolglos bleiben. Vom Standpunkt der empirisch-inhaltlichen Interpretationsabsicht muß daher für derartige Rekonstruktionsversuche von dispositionellen Eigenschaften gelten, was KALVERAM (1970) an faktorenanalytischen Klassifikationsverfahren kritisiert:

"Natürlich kann man auf eine inhaltlich vollständige Kategorisierung von Tests verzichten und die Kategorisierung lediglich hinsichtlich der gemeinsamen Faktoren vornehmen; hinsichtlich dessen, was sie erfassen, hat man die Tests nun allerdings nicht mehr klassifiziert, insofern man wichtigere Ordnungsgesichtspunkte außer acht gelassen hat. Ein so gewonnenes Ordnungsschema ist demnach ebenso willkürlich wie andere gleichdimensionale Schemata auch, eine Faktorenanalyse wäre in diesem Fall also eigentlich überflüssig gewesen" (ebd., S. 112).

Daß solche Ordnungsschemata willkürlich sind, hat seinen letzten Grund in der Tatsache, daß sie nicht den empirisch-genetischen Prozeß der Auseinandersetzung des untersuchten Individuums mit der diagnostischen Situation strukturieren, sondern reduzierte und isolierte Daten über die diagnostische Situation und über das Resultat der Auseinandersetzung - dem Verhaltensprodukt: Antwort auf die Testaufgabe - nach einem bestimmten, formal-methodischen Modell synthetisieren, ohne die Datengenesese selbst bei der Modellanwendung in Rechnung zu stellen. Vermutlich hat KALVERAM an Informationen aus dem Prozeß der Genese von Daten gedacht, wenn er für die vollständige Kategorisierung von Tests die Berücksichtigung von "wichtigeren Ordnungsgesichtspunkten" fordert. Schließlich lassen sich erst mit der Kenntnis der Genese von diagnostischen Daten jene Generalisierungen in Form von Urteilen über Eigenschaften von Personen begründen,

die sich im Sinne der unterstellten Bedeutung von Dispositionen als langfristig invariant erweisen sollen.

Mit der Kenntnis der Genese von diagnostischen Daten wäre zugleich der empirische Aspekt der Interpretationsmöglichkeiten von Eigenschaftsurteilen über Personen umrissen. Die Forderung der Untersuchung der empirischen Auseinandersetzung des untersuchten Individuums mit der diagnostischen Situation entspringt von daher nicht nur einem globalen Forschungsinteresse an einer präzisen und differenzierten Analyse von Phänomenen, sondern vor allem auch, im Hinblick auf den Bereich der Diagnostik im organisierten Lernen, der Legitimationsbedürftigkeit von Selektionsentscheidungen, die für den konkreten, betroffenen Schüler mit der Vergabe von konkreten Lebenschancen verbunden sind. Für die Bestimmung des intendierten Anwendungsbereiches pädagogisch-psychologischer Diagnostik, für seine theoretische Strukturierung und empirische Untersuchung bedarf es daher einer veränderten Ordnungs- und Orientierungsrahmens. Er muß die Forschung von allen Anfang an auf die Miteinbeziehung des Prozesses der Auseinandersetzung des untersuchten Individuums mit der diagnostischen Situation verpflichten, bzw. auf die Untersuchung des Interaktionssystems S_I festlegen. In diesem Ordnungs- und Orientierungsrahmen muß die objektiv zwischen den Interaktionssystemen S_I und S_F bestehende Wechselbeziehung aufgegriffen werden, um die bereits angesprochene naiv-empiristische, methodische Formalisierung von S_I zu vermeiden. Dabei ist davon auszugehen, daß die für die diagnostische Rekonstruktion von dispositionellen Eigenschaften konstitutiven Bedingungen, die verhaltensprovozierende Reizlage und die Resultate der Auseinandersetzung des untersuchten Individuums mit dieser Reizlage, aus der Perspektive der beiden Interaktionssysteme unterschiedlich interpretiert werden. Da es bei der diagnostischen Rekonstruktion von dispositionellen Eigenschaften letztenendes um die Begründung einer invarianten Beziehung zwischen Reizlage und Verhaltensprodukt geht und die Begründung von der Interpretation beider Bedingungen abhängt, letzteres aber - wie gesagt - zwischen beiden In-

teraktionssystemen differieren kann, muß an die Erforschung von Dispositionen folgende Forderung gestellt werden: Aufgabe der Diagnostik ist es, "die empirischen Merkmale R und V als Bestandteile von S_F so abzubilden, daß ihre Rekonstruktion in der Struktur und Funktion von R und V in S_I kongruent ist. Diese Forderung resultiert aus der Aufgabe von R und V, in S_F als Prämissen für den Schluß auf DE (dispositionelle Eigenschaft; d. Verf.) zu fungieren" (SCHMIDT, 1974, S. 23).

Sicherlich impliziert diese Forderung nur eine formale Bestimmung des Gegenstandsbereiches pädagogisch-psychologischer Diagnostik. Sie erlaubt aber, den intendierten Anwendungsbereich wissenschaftlich begründeter Beurteilungsprozesse in der Weise einzugrenzen, daß jetzt alle Formen der diagnostischen Rekonstruktion von dispositionellen Eigenschaften als scheinbar adäquate Forschungsansätze aus der rationalen Diskussion des Problemfeldes ausgeklammert werden können, die ihre Grundlagen lediglich aus dem Interaktionssystem S_F beziehen und deren Ergebnisse sich einer verallgemeinerbaren inhaltlichen Interpretation entziehen. Wenn man so will, stellt die Forderung ein Beispiel der "negativen Heuristik" im Sinne von LAKATOS dar, die dem Wissenschaftler sagt, welche Wege er zu vermeiden hat. In dieser Funktion soll der in der SCHMIDT'schen Forderung zusammengefaßte Orientierungsrahmen für die Untersuchung der Aussagefähigkeit der praktizierten Diagnostik in der Schule - der Zensurengebung und der Objektivierungstendenzen - dienen. Seine inhaltliche Auffüllung und Ausdifferenzierung ist in der gegenwärtigen Forschungssituation aus mehreren Gründen nicht möglich. Dazu ist zu berücksichtigen, daß die geforderte Kongruenzbeziehung zwischen S_I und S_F nicht vor der Untersuchung praktizierter diagnostischer Prozesse formal zu definieren ist, sondern nur eine Richtlinie einer Forschungsstrategie darstellt, die sich aus der derzeitigen Mängelanalyse schulischer Diagnostik ergibt. Wie noch genauer zu zeigen sein wird, basiert die gegenwärtig praktizierte Form der Schülerbeurteilung auf Reflexionsbemühungen, die ausschließlich innerhalb des Inter-

aktionssystems S_F stattfinden und genau deshalb innerhalb der Diagnostiker- und Forschergemeinschaft die Verständigungsschwierigkeiten in der Interpretation von Urteilen über Schüler hervorgerufen. Die Inkongruenz in der Rekonstruktion der empirischen Merkmale R und V innerhalb der Beurteilergemeinschaft ist also einer Folge der Inkongruenz ihrer Rekonstruktion mit jener, die jeweils schülerspezifisch, jedoch unhinterfragt im Interaktionssystem S_I stattfindet.

Deshalb kann die Rekonstruktion dispositioneller Eigenschaften nur das Resultat der Auseinandersetzung zwischen dem untersuchten Individuum und dem Diagnostiker mit R und V darstellen. Die erreichbare Kongruenz zwischen beiden Systemen ist somit nicht - wie es auf den ersten Blick erscheinen mag - durch ein formales Isomorphiemodell vor aller inhaltlichen Untersuchung des diagnostischen Prozesses festzulegen und zu programmieren. Sie ist vielmehr gleichzusetzen mit der Bedeutung, die beide Systeme durch Interpretation mit R und V verbinden und die zwischen beiden Systemen vermittelbar ist. Bedeutung muß dabei im weitesten Sinne des Wortes als übertragene Informationsinvariante gefaßt werden. Die erkenntnistheoretische Reflexion dieses Ansatzes ist m.E. bisher am weitesten von NARSKI (1968) in seiner "Theorie der funktionalen Bedeutung" vorangetrieben worden. Auf seine detaillierte Betrachtung muß hier verzichtet werden, obwohl die darin vertretene Auffassung gegenüber dem Problem der Dispositionsbegriffe für eine Neuorientierung der Schülerbeurteilung sehr fruchtbar sein dürfte.

Forschungsstrategisch impliziert dieser Ansatz eine völlig veränderte Konzeption der pädagogischen Diagnostik. Die Realisierung der Kongruenzforderung verweist auf die Einlösung von Wissenschaftlichkeitsprinzipien, die für die Handlungsforschung konstitutiv sind, (vgl. VAGT, 1972). Darin wäre auch und vor allem die Rekonstruktion der Schwierigkeiten und des defizitären Forschungszustandes zu leisten, die sich im Rahmen der derzeitigen Bestandsaufnahmen in der Bildungsforschung nur als Beschreibungen konkretisieren (vgl. HOPF, 1975).

2.3 Verfahrenstechnische Festlegung von Dispositionen durch die Zensurengebung

Selbst wenn man die Kongruenzforderung von SCHMIDT zunächst nur als negative Heuristik zur Erforschung des diagnostischen Prozesses begreift, ist zuerst inhaltlich zu klären, welche Wege der Wissenschaftler zu vermeiden hat. Da die Schülerbeurteilung immer an der Funktion des organisierten Lernens in seiner gesellschaftlich durchgesetzten Form orientiert ist und deshalb letztlich als Informationsgrundlage für Selektionsentscheidungen fungiert, muß die inhaltliche Klärung möglicher Irrtümer bei der diagnostischen Rekonstruktion von dispositionellen Eigenschaften an der schulsystembezogenen Beurteilergewohnheit ansetzen. Das allgemeinverbindliche Verfahren zur Schülerbeurteilung ist hier immer noch die Zensurengebung. Man sollte sich nicht damit zufriedengeben, die bekannte Kritisierbarkeit der Zensurengebung - ihre unzureichende Objektivität, Reliabilität und Validität - zur Kenntnis zu nehmen, ohne die Ursachen ihrer Schwächen hinterfragt zu haben. Erst die Untersuchung der tieferen Ursachen vermag dazu beitragen, daß der zu vermeidende Weg in seinen Einzelheiten erkannt wird und die Realisierung eines neuartigen Ansatzes als notwendig erscheint. An der folgenden idealtypischen Rekonstruktion der praktizierten Zensurengebung wird zu erkennen sein, wie die Kongruenzforderung unberücksichtigt bleibt und welche naiven Annahmen in die Legitimation der Beurteilung dispositioneller Eigenschaften eingehen. Auf der Basis dieser Betrachtung kann dann eine erste Bilanz hinsichtlich der beiden Thesen gezogen werden, daß die vorgenommene Formalisierung der Schülerbeurteilung durch die Zensurengebung keine Bestimmung des Anwendungsbereiches pädagogischer Diagnostik zuläßt und daß sich die Beurteilergemeinschaft nur über verfahrenstechnische Fragen des Beurteilungsprozesses unterhält.

Zu diesem Zweck gehe ich von einer Konkretisierung der diagnostischen Problemlage aus, wie sie mir für die praktizierte Beurteilung von Schülern im bestehenden organisierten Lernen .typisch zu sein scheint:

Als Exemplifizierung der diagnostischen Situation solle die Aufgabe für einen bestimmten Schülerjahrgang einen Erlebnisaufsatz mit einem bestimmten Thema zu schreiben, und die Durchführung der Aufgabe dienen. Da die Beurteilung des Aufsatzes durch den Lehrer ein Teilaspekt der übergreifenden Fähigkeits- und Persönlichkeitsbeurteilung der Schüler ist, steht der Lehrer bei der Aufsatzbeurteilung vor der Frage, nach welchen Kriterien der geschriebene Aufsatz so klassifiziert werden kann, daß er insgesamt als Ausdruck der Wirksamkeit von bestimmten Dispositionen bzw. Dispositionskomplexen zu begreifen ist. Damit wird gleichzeitig die Frage gestellt und zu beantworten versucht, ob die konkreten Operationen, die der Schüler zur Bewältigung der Aufgabe "Aufsatzschreiben" ausführte, transferierbar sind, d.h. in anderen, aber strukturell ähnlichen, diagnostischen oder sonstigen Bewährungssituationen wiederholt auftreten.

Setzt man voraus, daß es derartige, von situativen Kontexten relativ unabhängige und aktivierbare Operationen gibt, und diese auch grundsätzlich feststellbar sind, dann läßt sich daraus legitimerweise der Schluß ziehen, daß sie auch innerhalb der unmittelbar gegebenen, diagnostischen Situation vom Schüler reproduziert werden, wenn sich die gesamte diagnostische Situation dafür eignet. Damit ist die logische Voraussetzung für die Klassifizierbarkeit von geschriebenen Aufsätzen nach Merkmalen konstruiert, die mehr als einmal feststellbar sind und deshalb auf generative, dispositionelle Eigenschaften des Schülers verweisen.

Welcher Art können nun solche Kriterien bzw. Merkmale für die Klassifikation sein ?

Vorausgesetzt, der Beurteilungsprozeß ist funktional letztlich auf die Selektionsentscheidung ausgerichtet, dann muß der Lehrer zur Beurteilung des Aufsatzes verallgemeinerte Kriterien

heranziehen, d.h. Kriterien, die auch für andere Lehrer in der gleichen Beurteilungslage verbindlich sind. Das bedeutet wiederum, daß sich die pädagogischen Konsequenzen des Urteils - also die dem Schüler aufgrund des Urteils zugeordneten Selektionsalternativen - der Beeinflussung durch den beurteilenden Lehrer entziehen. Zur Veranschaulichung beziehe ich mich auf ein Beispiel aus der bayerischen Lehrplanreform; hier sieht der Stoffplan für den Bereich der Erlebniserzählung in der 6. Jahrgangsstufe vier Kriterien vor, die zugleich als lernzielartige Richtlinien für die Lehrplan- und Unterrichtsgestaltung fungieren: "klares Herausarbeiten der einzelnen Erlebnisschritte und des Höhepunktes, nähere Kennzeichnung der räumlichen und zeitlichen Umstände; Ersatz farbloser Wörter durch den treffenden Ausdruck, Vermeidung der Wiederholung, besonders gleicher Satzanfänge" (SCHULREFORM, 1970, S. 117).

Mit diesen vier Festlegungen erschöpft sich der Bereich der Klassifikationsmerkmale für die Beurteilung von Erlebnisaufsätzen. Der einzelne Lehrer steht nun vor der Aufgabe, den geschriebenen Aufsatz dahingehend zu untersuchen, ob und in welcher Häufigkeit Wörtern, Sätzen, Passagen, Abschnitten und dem Absatz als Ganzes die genannten Eigenschaften zugeordnet werden können. Er steht in einem zweiten Schritt vor der zusätzlichen Aufgabe, ein Modell der Synthese der Merkmalsausprägungen anzuwenden und zwar so, daß aus den Reproduktionsformen solcher Merkmale auf unterstellte, verursachende dispositionelle Eigenschaften geschlossen werden kann.

2.3.1 Das Dilemma in der Interpretation von Beurteilungskategorien

Die Durchführung der beiden Aufgaben, die Kennzeichnung des Aufsatzes durch Merkmale und die Synthese der Merkmalsausprägungen, beinhaltet zwei qualitativ verschiedene Prozesse. Im ersten Fall geht es um die Subsumption von selektierten Aspekte des Gesamt-

aufsatzes unter Kategorien, die als extern festgelegter Interpretationsrahmen an den Aufsatz herangetragen werden, also um die Klassifikation des vorliegenden A u f s a t z e s. Dies kann prinzipiell auch ohne die Absicht erfolgen, auf den Verfasser Rückschlüsse im Sinne seiner Verwendbarkeit für bestimmte Tätigkeiten oder seiner Geeignetheit für bestimmte Ausbildungszweige zu ziehen. Wird z.B. der Aufsatz als ein Kommunikationsmedium betrachtet, durch das der Verfasser seine Auseinandersetzung mit einem zu lösenden Problem darstellt, so kann die vorgeschlagene Problemlösung für einen Leser dahingehend wichtig sein, sich mit dem Verfasser darüber eingehender zu unterhalten und den Lösungsprozeß zu vervollständigen und nicht, sich für den Verfasser als potentiellen Reproduzenten gleicher Produkte unter veränderten situativen Bedingungen zu interessieren. Diese Bemerkung ist im Hinblick auf die Frage wesentlich, wie aus einer Klasse von vergegenständlichten Verhaltensweisen in Relation zu einer Klasse von festgelegten, diagnostischen Bedingungen auf zugrundeliegende dispositionelle Eigenschaften zu schließen ist. Sie ist dann wesentlich, wenn der Gesichtspunkt der Klassifikation von der Verwendung oder Weiterentwicklung des Verhaltensresultates her bezogen wird und nicht aus dem empirischen Konstitutionsprozeß der Verhaltensresultate - obwohl daraus die dispositionelle Eigenschaft begründet werden soll.

Den vorher zitierten vier Kriterien,

1. Klares Herausarbeiten der einzelnen Erlebnisschritte und des Höhepunktes,
 2. nähere Kennzeichnung der räumlichen und zeitlichen Umstände,
 3. Ersatz farbloser Wörter durch den treffenden Ausdruck,
 4. Vermeidung der Wiederholung, besonders gleicher Satzanfänge,
- ist unmittelbar nicht anzusehen, ob sie nur empirisch unbegründbare Diskriminationskriterien für Schüleraufsätze, also für vergegenständlichte Verhaltensresultate sind oder auch Merkmale von Schüleraufsätzen, die auf empirische Konstitutionsoperationen dieser Aufsätze verweisen.

Zweifellos stellen diese Kriterien Konventionen dar, die sich in der praktizierten Beurteilung von Schüleraufsätzen für deren Unterscheidung bewährt und die für Selektionszwecke partiell gute Dienste erbracht haben. Ordnet man jedem unterscheidbaren Aufsatz seinen Verfasser zu und teilt die Schüler nach diesem Verfahren in Niveaugruppen ein, dann besitzen Aufsätze als Selektionsverfahren im bestehenden Schulsystem in pragmatischen Grenzen und regional beschränkten, schulischen Verwendungsbezirken eine erstaunlich hohe Vorhersagegültigkeit, jedenfalls solange, wie man den Bedeutungsgehalt solcher Urteile nicht untersucht. Die Interpretationsbeliebigkeit von Urteilen über Schüler auf der Basis von Aufsätzen tritt ja erst dort eklatant in Erscheinung, wo die Verwendbarkeit des Aufsatzes als Selektionsverfahren zu verallgemeinern versucht wird, sei es über die Beurteilerpopulation wie z.B. in der Untersuchung von SCHRÖTER (1971), die Anwendungspopulation, über die regionalen Spezifitäten von Schulen, die Anforderungsunterschiede von aufnehmenden Schulen usw. Der Erforschung dieses Phänomens, also der objektiven Bedingungen für die Möglichkeit einer relativen, prognostischen Gültigkeit des traditionellen Lehrerurteils, hat sich die empiristisch orientierte Forschung bisher weitgehend entzogen. Dies wäre aber gerade im Hinblick auf die Erwartung notwendig, durch die Objektivierung der Schülerbeurteilung nicht nur präzisere, sondern vor allem inhaltlich interpretierbare Voraussagen des Schulerfolgs zu erreichen, die über die prognostische Gültigkeit des Lehrerurteils hinausgeht.

Ein weiteres Kennzeichen der vier Kriterien hängt mit der historisch nachweisbaren Tatsache zusammen, daß sie globalisierte, literarische Kategorien zur Beurteilung von Texten sind. Sie können in ihrer Formalität prinzipiell auf jede Art von Text Anwendung finden. Zwar haben solche Kriterien eine inhaltliche Bezugsgrundlage, auf der sie entwickelt wurden und von der sie letztlich nicht gelöst werden könnten. Es ist z.B. bereits intuitiv einsichtig, daß das Kriterium "Klares Herausarbeiten der einzelnen Erlebnisschritte und des Höhepunktes" nicht ohne wei-

teres auf eine Sachbeschreibung anzuwenden ist. Die Aufsatzkategorie "Erlebniserzählung" ist aber so breit angelegt, daß die Einlösung und Anwendung der vier Kriterien unabhängig von der inhaltlichen Auseinandersetzung durchzuführen ist. Auf jeden Fall sagen sie über die empirischen Prozesse der Auseinandersetzung des Aufsatzverfassers mit dem Thema nichts aus. Bleibt die inhaltliche Bezogenheit und gegenseitige Abhängigkeit der Kriterien bei der Aufsatzbeurteilung dadurch unberücksichtigt, daß sie als einzige allgemeinverbindliche Beurteilungsorientierungsgrundlage wie in diesem Beispiel Gültigkeit besitzen, so besteht ihre Funktion lediglich darin, die **F o r m** **d e r D a r s t e l l u n g** der inhaltlichen Auseinandersetzung zu klassifizieren. Berücksichtigt man außerdem den Umstand, daß diese Kriterien zugleich als lernzielartige Richtlinien für die Lehrinhaltsvermittlung wirksam sind und die Schüler im Verlauf des Unterrichts dazu angehalten werden, bei der Aufsatzgestaltung diese "kritischen" Forderungen zu erfüllen, dann liefert der Unterricht die Vorbedingung dafür, daß sie im nachhinein Anwendung bei der Aufsatzbeurteilung finden können. Damit wird die Voraussetzung geschaffen, die Kriterien, losgelöst von der inhaltlichen Auseinandersetzung des Schülers mit dem im Thema umrissenen Gegenstand und isoliert aus ihrem gegenseitigen Zusammenhang, als Beurteilungsraster an den Aufsatz anzulegen. Dem Ergebnis der konkreten Subsumption des komplexen Aufsatzes unter die Kriterien durch den beurteilenden Lehrer haftet damit jene Beliebigkeit an, die in der feststellbaren intersubjektiven Inkonsistenz des Lehrerurteils in Erscheinung tritt, während die dennoch vorhandene, in Grenzen gegebene prognostische Gültigkeit des Lehrerurteils unverstanden bleibt.

Für die Untersuchung der Frage, ob sich der Gesamtprozeß der Auseinandersetzung des untersuchten Individuums mit der diagnostischen Situation - hier der Aufforderung, einen Aufsatz mit einem bestimmten Thema zu schreiben - unter eine vorgegebene und vorausgesetzte Bedeutung von Klassifikationskriterien der Reizlage und der Verhaltensresultate subsumieren läßt, bzw.

aus diesen miterklärt werden kann, ist zunächst einmal festzustellen, daß die vier genannten Kriterien diese Bedingung nicht erfüllen. Das liegt zum einen daran, daß die Bedeutung der Klassifikationskriterien nicht aus der inhaltlichen Auseinandersetzung des untersuchten Individuums mit der diagnostischen Situation als verallgemeinerbares Kennzeichen der Auseinandersetzung gewonnen wurde und deshalb in Bezug darauf zwischen den Beurteilern keine Vermittlung möglich ist. Die Interpretation einer bestimmten Menge von Aufsatzbestandteilen unter den kategorialen Gesichtspunkten als Ausdruck von dispositionellen Eigenschaften muß der privaten Auffassung des einzelnen Beurteilers überlassen bleiben und ist daher subjektivistisch. Die empiristische Sozialpsychologie hat sich die Untersuchung der subjektivistischen Konstitution von Urteilen über andere auf der Basis vorausgesetzter Klassifikationsgesichtspunkte des Verhaltens und der das Verhalten provozierenden Situation zum Gegenstand gemacht. Sie muß sich in Ermangelung einer empirischen Begründbarkeit der Klassifikationsgesichtspunkte mit der Entwicklung von infiniten Variablenmodellen zufriedengeben.

Zum anderen sind die Erklärungsschwierigkeiten dadurch zu begründen, daß die ursprünglich inhaltsgebundenen Klassifikationsgesichtspunkte von Texten durch ihre Formalisierung als voneinander unabhängige Kriterien in ihrem eigenen Konstitutionszusammenhang nicht mehr aufgeklärt werden können und damit ihre eigentliche Bedeutung als Formbestimmungen der Darstellung von sprachlich-schriftlichen Auseinandersetzungen mit einem Gegenstand nicht mehr intersubjektiv konsistent vermittelt werden kann. Anschaulich gesprochen: Der Beurteiler muß erst noch einmal innerhalb des Spielraums gesellschaftlich geltender Normen für sich selbst interpretieren, was er eigentlich unter "klarem Herausarbeiten der einzelnen Erlebnisschritte und des Höhepunktes" usw. verstehen will. Die Lösung dieser Frage ist z.B. ein zentrales Problem der Lernzielbestimmung, der sog. Lernzieloperationalisierung, der Curriculumentwicklung, der didaktischen Forschung etc. Es wird dort der Versuch unternommen, den Sub-

jektivismus des Interpretationsprozesses durch zunehmende Planung der Ausbildung zu reduzieren und Intersubjektivität in der Verwendung von Klassifikationsgesichtspunkten der Beurteilung durch die Herstellung von Konsensus auf hierarchisch gegliederten und institutionalisierten Ebenen zu vermitteln.

Die empiristische Rekonstruktion der Bedeutung der auf diese Weise intersubjektiv vermittelten Klassifikationsgesichtspunkte im Bereich pädagogisch-psychologischer Forschung ist allerdings bis jetzt erfolglos geblieben. Ganz allgemein ist festzustellen, daß entgegen der ursprünglichen Erwartungen, es bis heute nicht gelungen ist, Lernziele und Lernzielstrukturen weder durch phänomenologische Analysen und Konsensustechniken noch durch empiristische Rekonstruktion zu objektivieren, d.h. als intersubjektiv vermittelbar zu konkretisieren. So erbrachte z.B. der Versuch, die Taxonomie von BLOOM zu validieren, nicht die erwarteten Erfolge (KROPP & STOKER, 1966).

Abgesehen von den Verfahrensschwierigkeiten, die ein derartig komplexer Validierungsversuch impliziert, ist man sich in der Grundlagenforschung über die methodologischen Probleme in der Rekonstruktion der intersubjektiv vermittelbaren Bedeutungsgehalte von Urteilen, Klassifikationen bzw. Begriffen in sozialpsychologischen Kontexten in weiten Bereichen nicht bewußt. Innerhalb der empiristischen Denkweise wäre für die Herstellung eines intersubjektiven Konsensus im Gebrauch von Urteilen bzw. Klassifikationen, über die per Voraussetzung noch kein Konsens besteht, eine intersubjektiv gebrauchte Metasprache zur Vermittlung des allgemeinen Bedeutungsgehaltes solcher Urteile bzw. Klassifikationen notwendig; um dem sich hier andeutenden Regreß zu entgehen, versucht man, die Vermittelbarkeit durch Methoden und Techniken zu reglementieren.

Durch die Methode des "semantischen Differentials" (OSGOOD, SUCI & TANNENBAUM, 1957) wird der Bedeutungsgehalt von Begriffen durch Dimensionen eines semantischen Raumes strukturiert (Valenz-Dimension, Potenz-Dimension, Aktivitätsdimension), die auf faktorenanalytischem Weg ermittelt werden. Das entscheidende Pro-

blem dabei ist, daß die ermittelten Dimensionen auf der Einschätzung der Position jener Begriffe in semantischen Feldern basieren; die methodisch reglementierte Ermittlung der Begriffsbedeutung hat also eine Vorbestimmung der resultierenden Begriffsbedeutung zu ihrer Voraussetzung. Dies ist ein partiell zirkuläres Vorgehen; partiell deshalb, weil die Vorbestimmung auf die vorgegebenen semantischen Felder zurückgreift. Letztere bestehen aus einer Reihe von bipolaren Adjektivpaaren wie heiß-kalt, groß-klein, angenehm-unangenehm u.ä. Für die Einschätzung der Position der zu untersuchenden Begriffe in diesen Feldern muß nun wieder die stillschweigende Voraussetzung gemacht werden, daß über die Bedeutung der Adjektive sowie über die zwischen den Paaren angenommene Polarität intersubjektiver Konsens besteht. Diese fragwürdige Annahme ließe sich nicht nur empiristisch widerlegen durch die Ermittlung des semantischen Differentials der Adjektive (von dem methodischen Problem, daß das untersuchte Adjektivpaar dann in der Reihe der zugrundegelegten Adjektivpaare nicht mehr vorhanden wäre, könnte man hier absehen). Sie impliziert vielmehr ein mechanistisches Modell der Genese von Begriffsbedeutungen, nach dem sich die Bedeutung von komplexen, intersubjektiv teilweise inkonsistent gebrauchten Begriffen auf einfachere, intersubjektiv konsistent gebrauchte zurückführen und in entsprechen umgekehrter Reihenfolge rekonstruieren läßt.

Das Modell führt in letzter Konsequenz zu der Unterstellung, intersubjektiv vermittelte Begriffsbedeutungen ließen sich kontextfrei bestimmen. Auf die Unhaltbarkeit dieser These kann hier nicht eingegangen werden. Sie widerspricht sowohl dem Forschungsstand der strukturalen Linguistik (vgl. BECHERT u.a., 1971), als auch den Ergebnissen der wissenschaftlichen Untersuchung des Spracherwerbs (vgl. OLDFIELD & MARSHALL, 1968). Im besonderen ist aber auf die Illusion hinzuweisen, die sich mit der Erwartung verbindet, die intersubjektive Bedeutung von Begriffen durch methodische Objektivierung gewinnen zu wollen. Sie ist für die Methode des semantischen Differentials

konstitutiv. Vermittelbare Bedeutung wird mit wiederholbaren statistischen Korrelationen gleichgesetzt. Die tautologische Struktur dieses Vorgehens hat FODOR (1965) in seinem Aufsatz über "Could meaning be an r_m ?" differenziert aufgezeigt. Andere, im Prinzip ähnliche und vor allem unter demselben kritischen Gesichtspunkt zu problematisierende Methoden und Techniken zur Objektivierung des Bedeutungsgehaltes von Klassifikationen bzw. Urteilen wie etwa die Q-Technik von STEPHENSON bedürfen keiner besonderen Erörterung, da im Vordergrund der Überlegungen das Verhältnis von Begründbarkeit von Klassifikationen und Funktionalität von methodisch formalisierten Untersuchungsansätzen steht (vgl. die zusammenfassende Darstellung und differenzierte Kritik des semantischen Differentials und der Q-Technik von MEYER, 1972). Man begegnet auch hier wieder der gleichen Schwierigkeit, die in der Behandlung der Objektivitätsfrage sowie in der Diskussion der Faktorenanalyse für die Rekonstruktion von dispositionellen Eigenschaften aufgetaucht ist, nämlich der Schwierigkeit, daß die methodische Reglementierung wissenschaftlicher Untersuchungen keinerlei Information über das begriffliche Verständnis des untersuchten Gegenstandes erbringt.

Es zeigt sich bei den soeben angesprochenen Formen der empiristischen Rekonstruktion des Bedeutungsgehaltes von Klassifikationen, daß sie grundsätzlich nur die in der Beurteilerpopulation wirksamen, zentralen, normorientierten Tendenzen in der Interpretation von Klassifikationsgesichtspunkten quantitativ beschreiben. Die Möglichkeit zur Feststellung wird dabei erst dadurch geschaffen, daß die Komplexität des Bedeutungsfeldes von Klassifikationsgesichtspunkten durch ein quantifizierendes Untersuchungsmodell des Forschers vorstrukturiert wird. Man kann daher derartige Objektivierungsverfahren durch die Beurteilung charakterisieren, die MEYER für den Stellenwert der Q-Technik abgibt, wenn man gleichzeitig die starke und bei ihm unproblematisierte Forderung nach nomologischen Hypothesen durch die Forderung nach einer begrifflichen Gegenstandsana-

lyse ersetzt:

"Der Einsatz der Q-Technik ist ... erst dann lohnend, wenn mehrere nomologische Hypothesen über die Beziehungen zwischen den einzelnen Variablen des zu analysierenden Problems bereits entwickelt worden sind. Die 'bloße' Ermittlung der Normpräferenzen bestimmter Personen kann dagegen auch mit einfacheren Erhebungsmethoden, z.B. mit Fragebogen und Interviews durchgeführt werden. Es wäre also eine unberechtigte Hoffnung, daß die Qualität des Forschungsinstruments die Dürftigkeit der in der Disziplin vorliegenden empirisch gehaltvollen Hypothesen überspielen könnte; ganz im Gegenteil: je anspruchsvoller das Meßinstrument, umso offensichtlicher wird auch das Fehlen von Hypothesen" (ebd., S. 177).

Die erwähnte Methodenproblematik, die vor allem im Rahmen der Modelle der "Judgement Analysis" (vgl. EDWARDS, 1967; KLEINMUNTZ, 1968; NEWELL & SIMON, 1965) virulent wird, weil dort durch die laborexperimentelle und simulationstechnologische Instrumentalisierung der Bedeutungsfindung bei Urteilen die Begründung und Legitimation von Klassifikationen und hypothetischen Aussagen eine untergeordnete Rolle spielt, wäre auch an den entscheidungstheoretischen Ansätzen und Konsensustechniken in der Untersuchung der Urteilsbildung aufzuweisen. Auch hier bleibt die Frage nach der Aufklärbarkeit des Bedeutungsgehaltes von eingeführten Klassifikationsgesichtspunkten zur Beurteilung von vergegenständlichten Verhaltensresultaten unbeantwortet und wird durch die Reglementierung des Entscheidungsprozesses bzw. der Konsensusfindung kompensiert (FLECHSIG u.a., 1970; HELMER, 1966). Es erübrigt sich aber, darauf einzugehen, weil die Ausgangssituation für die Rekonstruktion von dispositionellen Eigenschaften auf der Grundlage von inhaltlich ungeklärten Klassifikationsgesichtspunkten zur Unterscheidung von Verhaltensresultaten die gleiche bleibt.

Die Ausgangssituation gilt es für die Fortsetzung der Betrachtung des Veranschaulichungsbeispiels systematisch zusammenzufassen. Vereinfacht ausgedrückt ging es im ersten Schritt um die Beurteilung des Aufsatzes und nicht um die Beurteilung des Schülers als Verfasser des Aufsatzes. Die Diskussion der intersubjektiv konsistenten Interpretierbarkeit der

Klassifikationsgesichtspunkte bewegte sich deshalb ausschließ-
lich im Bezugsrahmen jenes Interaktionsgeschehens, welches sich
innerhalb der Diagnostiker- bzw. Forschergemeinschaft vollzieht.
Die sich in der Beschränkung auf dieses Interaktionssystem
konkretisierende Verselbständigung eines Teilprozesses der dia-
gnostischen Rekonstruktion von dispositionellen Eigenschaften
ist zwar vor dem Hintergrund der SCHMIDTschen Forderung als naiv-
empiristisch zu kritisieren; sie war aber notwendig, weil das
Vorgehen der praktizierten Diagnostik in der Schule untersucht
werden soll und hier das Interaktionsgeschehen im Interaktions-
system S_F in verselbständigter Form organisiert ist.

Allgemein kann man als Ergebnis der Betrachtung festhalten, daß
die Notwendigkeit zur subjektivistischen Interpretation für
vergegenständlichte Verhaltensresultate - im vorliegenden Bei-
spiel der Aufsatz - so, wie sie dem beurteilenden Lehrer auf-
getragen ist, einerseits die Objektivierung des Interaktions-
prozesses erforderlich macht. Auf der anderen Seite bleiben aber
die Objektivierungsmodelle, seien es Konsensustechniken, ent-
scheidungstheoretische Ansätze oder Versuche der empiristischen
Rekonstruktion des vermittelbaren Bedeutungsgehaltes von Klas-
sifikationsgesichtspunkten, erfolglos, weil sie selbst überge-
ordnete und intersubjektiv verständliche Klassifikationsgesichts-
punkte für die Objektivierung des ersteren voraussetzen müssen.
Für die Probleme der Lernzielfindung und -operationalisierung,
der Curriculumplanung und -evaluation, der Qualifikationsfest-
legung und -beurteilung folgt daraus, daß in diese Prozesse
subjektivistische Einflüsse auf den verschiedenen Ebenen ihrer
Organisation und Institutionalisierung eingehen. In dem Maße
wie die subjektivistischen Einflüsse auf den verschiedensten
Ebenen wirken und gleichzeitig einer Erklärung und Legitimation
entzogen werden, geraten die Objektivierungsmodelle zu formalen
Ordnungsschemata, die zwar zur Reglementierung und Standardisie-
rung des gesamten Ausbildungssystems beitragen, die reale Durch-
setzung der Regeln und Standards aber der privaten Initiative
und Kompetenz des einzelnen Lehrers überlassen.

2.3.2 Formale Synthese dispositioneller Eigenschaften: die Funktionalisierung des Anwendungsbereiches pädagogischer Diagnostik

Auf welche Weise kann der einzelne Lehrer angesichts der Interpretationsspielräume in den curricular vorgeschriebenen Klassifikationsgesichtspunkten zur Beurteilung von Verhaltensresultaten zur diagnostischen Rekonstruktion dispositioneller Eigenschaften von Schülern gelangen ?

Er befindet sich unter den Bedingungen seiner letztlich selektionsfunktionalen Beurteilertätigkeit in der widersprüchlichen Situation, die verschiedenen Aufsätze unter ihm vorgegebene, formale Kategorien - die vier oben genannten Kriterien - subsumieren zu müssen. Das heißt, er hat die formalen Kategorien mit Hilfe der Aufsätze entsprechend seinen Vorstellungen, Kompetenzen und Erfahrungen zu interpretieren und daraus Beurteilungen der Aufsätze zu konstruieren, die grundsätzlich für andere Lehrer vermittelbar sein sollen, oder anders gesagt, die mit den Beurteilungen anderer Lehrer vergleichbar sind.

Die Rekonstruktion des widersprüchlichen Beurteilungsprozesses wird zeigen, daß die Vergleichbarkeit der Urteile nur um den Preis ihrer inhaltlichen Nichtinterpretierbarkeit aufrechtzuhalten ist. Sie führt zu einer Formalisierung des Interaktionsgeschehens in S_F , die eine Bezugnahme auf das Interaktionsgeschehen in S_I sogar auf empiristischem Wege unmöglich macht. Zum Zweck der Rekonstruktion ist eine Unterscheidung in Bezug auf den Verwendungskontext der Aufsatzbeurteilung durchzuführen: Im ersten Fall soll zunächst davon abgesehen werden, daß die Aufsatzbeurteilung konstitutiver Bestandteil der diagnostischen Informationsgrundlagen für Selektionsentscheidungen im weitesten Sinn des Wortes ist. Der einzelne Lehrer möge zunächst den Gesichtspunkt außerachtlassen, seine Beurteilungen müßten mit jenen anderer Lehrer vergleichbar sein. Im zweiten Fall wird dieser Gesichtspunkt wieder einogenolt und die Widersprüchlichkeit des Beurteilungsprozesses konkretisiert.

Insgesamt wird mit der Rekonstruktion des Beurteilungsprozesses die Absicht verfolgt, jenen Aspekt der praktizierten Diagnostik in der Schule darzustellen, der sich überhaupt verallgemeinern läßt - die formale Synthese von subjektivistisch interpretierten Merkmalen vergegenständlichter Verhaltensresultate in Prüfungssituationen.

Rekonstruktionsmodell I

Im ersten Rekonstruktionsmodell wird davon ausgegangen, daß der Lehrer die Kriterien als **k l a s s i f i k a t o r i - s c h e** Prädikate mit zwei alternativen Interpretationen verwendet und die Anwendung der verschiedenen Kriterien als unabhängig voneinander betrachtet. **M.a.W.:** Die Kriterien sind als Merkmale des Aufsatzes mit jeweils zwei Ausprägungen anzusehen, deren Gegebenheit unabhängig voneinander festzustellen ist. Übertragen auf die exemplarischen Beurteilungskriterien für den Aufsatz, ergeben sich daraus Beurteilungen der folgenden Art:

- einzelne Erlebnisschritte und der Höhepunkt sind klar herausgearbeitet oder sie sind es nicht;
- die räumlichen oder zeitlichen Umstände sind näher gekennzeichnet oder sie sind es nicht;
- farblose Wörter sind durch den treffenden Ausdruck ersetzt oder sie sind es nicht;
- Wiederholungen, besonders gleicher Satzanfänge sind vermieden oder sie sind es nicht,

Die im ersten Rekonstruktionsmodell gemachten Voraussetzungen sind keineswegs selbstverständlich, sie sind aber für den Übergang zum zweiten Modell mit Ausnahme der Dichotomisierung "Merkmal vorhanden oder nicht vorhanden", die nur zur Vereinfachung der Darstellung durchgeführt wird, erforderlich. Der Verzicht auf die Horeinnahme von komparativen bzw. metrisierten Prädikaten ist durch den Umstand zu begründen, daß die damit verbundene Problematik bis jetzt noch nicht erörtert wurde, vor allem aber durch den Umstand, daß sie bei der Durchführung der praktizierten Diagnostik in der Schule übergangen wird.

Da der Übergang zu scheinquantitativen Begriffen gerade im zweiten Rekonstruktionsmodell nachgezeichnet werden soll, ist dies ein weiterer Grund für die Beschränkung auf die Betrachtung von klassifikatorischen Prädikaten. Was die Voraussetzung der Unabhängigkeit der Kriterien anbelangt, so wurde bei der Diskussion der empiristischen Rekonstruierbarkeit von Bedeutungsgelalten bereits auf ihre Unhaltbarkeit hingewiesen. Auch sie wird wiederum im Hinblick auf die Darstellung der formalen Synthese von Merkmalsausprägungen in der praktizierten Diagnostik erforderlich, weil diese als lineare Kombination von voneinander unabhängigen Größen rekonstruiert werden muß.

Darüberhinaus ist auch eine pragmatische Voraussetzung zu machen: Man muß davon ausgehen, daß der Lehrer trotz aller möglichen kritischen Einwände eine Festlegung der Beurteilung von Aufsätzen in der angegebenen Art, "vorhanden oder nicht vorhanden", entweder vorher oder im Vollzug des Beurteilungsprozesses für sich vorgenommen hat und die Ergebnisse der Klassifikation vorliegen.

Auf dieser Basis und im Zusammenhang mit den gemachten Voraussetzungen kann er die verschiedenen Aufsätze nach Mustern klassifizieren, die sich als eine Folge von festgestellten Merkmalsausprägungen darstellen lassen. Ordnet man den verschiedenen Kriterien die Buchstaben A, B, C und D, und der Merkmalsausprägung "Vorhanden" eine 1 bzw. der Merkmalsausprägung "Nicht vorhanden" eine 0 zu, dann lassen sich die verschiedenen Muster als Variationen von zwei Elementen (zwei Merkmalsausprägungen) zur vierten Klasse (vier Kriterien) symbolisieren. Ihre Gesamtanzahl beträgt 16 ($= 2^4$).

Die folgende Tabelle stellt die verschiedenen Beurteilungsmuster im Überblick dar. Die letzte Spalte mit dem Titel "Häufigkeit der Merkmalsausprägung 'vorhanden'" wird erst bei der Untersuchung des zweiten Rekonstruktionsmodells zur Sprache kommen; sie hier anzuführen ist notwendig, weil dort auf die gesamte Tabelle zurückgegriffen werden muß.

Kennziffer des Musters	Kriterien				Häufigkeit der Merkmals- ausprägung "vorhanden"
	A	B	C	D	
1	1	1	1	1	4
2	1	1	1	0	3
3	1	1	0	1	3
4	1	0	1	1	3
5	0	1	1	1	3
6	1	1	0	0	2
7	1	0	0	1	2
8	0	0	1	1	2
9	1	0	1	0	2
10	0	1	0	1	2
11	0	1	1	0	2
12	1	0	0	0	1
13	0	1	0	0	1
14	0	0	1	0	1
15	0	0	0	1	1
16	0	0	0	0	0

Tabelle 1: Zusammenstellung der Beurteilungsmuster für vier Klassifikationsgesichtspunkte mit je zwei Interpretationen ("vorhanden" und "nicht vorhanden").

Mit dieser schematischen Darstellung ist nur die Symbolisierung der Ergebnisse einer Aufsatzbeurteilung erreicht, deren allgemeinverbindliche Festlegungen - jene Festlegungen, die von Lehrer zu Lehrer vermittelbar sind -, sich in den Bestimmungen des zitierten Beispiels aus der Lehrplanreform erschöpfen. Zur Veranschaulichung seien einige Muster durch die Festlegungen interpretiert.

Muster 1 besagt z.B., daß in dem entsprechenden Aufsatz alle geforderten Kriterien erfüllt sind, d.h., daß die Merkmalsausprägung "vorhanden" viermal konstatiert werden konnte. Muster 16 besagt das genaue Gegenteil, während in Muster 7 zum Ausdruck kommt, daß in dem betreffenden Aufsatz die einzelnen Erlebnis-

schritte und der Höhepunkt herausgearbeitet und Wiederholungen, besonders gleicher Satzanfänge, vermieden wurden, wohingegen von einer näheren Kennzeichnung der räumlichen und zeitlichen Umstände sowie vom Ersatz farbloser Wörter durch den treffenden Ausdruck nicht die Rede sein kann.

Wir nehmen an, der Lehrer würde bereits in diesem Stadium den Prozeß der Aufsatzbeurteilung abschließen und mit Hilfe der vorliegenden Interpretation des Aufsatzes zur Beurteilung seines Verfassers übergehen wollen. Er würde damit unbeschadet gegen alle kritischen Einwände gegen die Aufklärbarkeit des Bedeutungsgehaltes solcher Urteile über Texte mit der diagnostischen Rekonstruktion von dispositionellen Eigenschaften beginnen. Die Einwände ergeben sich aus dem intersubjektiv inkonsistenten Gebrauch der Klassifikationsgesichtspunkte im Interaktionssystem S_F und aus der Unausgewiesenheit des empirischen Bedeutungsgehaltes der Klassifikationsgesichtspunkte im Interaktionssystem S_I .

Da gemäß den unterstellten Voraussetzungen die Kriterien als klassifikatorische Begriffe eingeführt wurden und voneinander unabhängig sind, und mit der Unabhängigkeitsforderung auch die getrennte Anwendung der Kriterien durchgesetzt ist, darf er die Ziffern 1 und 0 nicht als Abstufungen von Merkmalsausprägungen interpretieren und die vier qualitativ verschiedenen Merkmalsausprägungen nicht als Resultat der Wirksamkeit von *u m f a s s e n d e n* oder *z e n t r a l e n* Dispositionen auffassen. Die Verknüpfung der symbolisierenden Ziffern 1 und 0 durch mathematische Operationen innerhalb eines Musters ist damit ebenso gegenindiziert wie ein darauf aufbauender Vergleich zwischen verschiedenen Mustern, also zwischen verschiedenen Aufsätzen.

Da mit der Eignungsbeurteilung von Schülern das Ziel verfolgt wird, beurteilte Verhaltensweisen auf die Wirksamkeit von Dispositionen zurückzuführen, müssen die verschiedenen, konstatierbaren Verhaltensmuster durch die Vorhandenheit bzw. Nichtvorhandenheit von unabhängigen, dispositionellen Eigenschaften im Schüler erklärt werden. Die damit vollzogene, funktionale

Zuordnung von Schülern zu Aufsätzen würde die Unterscheidung von maximal 16 Schülertypen auf der Grundlage von maximal 16 Aufsatzbeurteilungsmustern erlauben.

Der Prozeß der Zuordnung ist Ausdruck eines naiven Modells der diagnostischen Rekonstruktion von dispositionellen Eigenschaften. Die Naivität des Vorgehens tritt hier nur deshalb so deutlich in Erscheinung, weil die realiter wirksamen und zusätzlichen, nicht reglementierten Beurteilungsoperationen des einzelnen Lehrers außerhalb der Betrachtung bleiben und keine formalmethodische Verarbeitung der Beurteilungsdaten vorgenommen wurde. Das Modell impliziert das Verfahren, das Verhalten des Individuums bzw. die Erfassung des Verhaltens als Funktion von außerhalb des Individuums liegenden Bedingungen von Verhalten zu begreifen. Dabei werden die Verhaltensbedingungen nachträglich, d.h. erst nachdem der geschriebene Aufsatz vorliegt, als Interpretation von Klassifikationsgesichtspunkten konkretisiert; sie wurden jedoch auf der anderen Seite vor der Prüfungssituation als lernzielartige Richtlinien für die Lehrinhaltsvermittlung im ablaufenden Unterricht in anderer Weise realisiert. Man denke zum Beispiel an die Tatsache, daß im Unterricht das Aufsatzgestaltungsprinzip "Klares Herausarbeiten der einzelnen Erlebnisschritte" anhand eines bestimmten literarischen Textes veranschaulicht wurde, bei der Aufsatzbeurteilung aber diese Veranschaulichungsform keine Rolle mehr spielt, sondern mit Hilfe der vorliegenden, konkreten Aufsätze zu reinterpretieren versucht wird.

Mit der Konkretisierung von Klassifikationsgesichtspunkten durch ad-hoc-Interpretationen und mit der formalen Zuordnung von dadurch erhaltenen Beurteilungsmustern zu den Verfassern der Aufsätze vollzieht sich defacto eine vermögenspsychologische Festlegung bzw. Definition von instrumentalisierten, dispositionellen Eigenschaften. Durch diese Art der Beurteilung von Aufsätzen und durch die Ausführung des Zuordnungsprozesses werden Schüler als potentielle Träger von vier qualitativ verschiedenen Dispositionen gesehen, und jenachdem, ob die Merkmalsausprägungen als vorhanden feststellbar sind oder nicht, als Trä-

ger von verschiedenen Dispositionsmustern beurteilt. Auch ohne Berücksichtigung realiter stattfindender Beurteilungsvorgänge, wie etwa im Fach Deutsch die Beurteilung der Beherrschung grammatikalischer Regeln oder der Rechtschreibung und von Qualifikationen in anderen Fächern, müßte daraus eine pädagogische Konsequenz gezogen werden, die sich mit dem bestehenden, selektionsorientierten Schulsystem nicht vereinbaren läßt, und die sich prinzipiell mit keinem Ausbildungssystem vereinbaren läßt, das als Aufeinanderfolge von institutionalisierten Ausbildungsalternativen organisiert ist: Für jedes Beurteilungsmuster müßte eine spezifische und organisierte Ausbildungsalternative geschaffen werden, wenn das Erreichen der Kriterien zum Fundamentum der Ausbildungsaufgaben des Schulsystems gehört. Dies folgt aus den Voraussetzungen des Beurteilungsvorgehens, wonach die Beurteilungskriterien als klassifikatorische und voneinander unabhängige Prädikate eingeführt wurden.

Wären im Gegensatz dazu die Klassifikationskriterien als komparative bzw. metrische Begriffe, die durch gesetzesartige Annahmen zueinander in Beziehung gestellt werden könnten, in die Beurteilersprache eingeführt, dann könnte man wenigstens partiell auf die Organisation spezifischer Ausbildungsalternativen verzichten; partiell insoweit, wie die Merkmale als homogen gelten können und sich durch ihre methodisch zulässige Kombination Mengen von Aufsätzen mit gleichen Gesamtausprägungsgraden des unterstellten dispositionellen Faktors gruppieren ließen. Denn unter dieser Voraussetzung könnte man das vergegenständlichte Verhaltensresultat "Aufsatz" unter dem Gesichtspunkt graduell abgestufter Merkmalsausprägungen beurteilen, für die ein gemeinsamer, verursachender dispositioneller Faktor angenommen werden kann. Die Darstellung des zweiten Rekonstruktionsmodells wird zeigen, daß in der praktizierten Diagnostik in der Schule das Kombinationsverfahren auch ohne Klärung der Begriffsproblematik zur Anwendung kommt.

Es liegt auf der Hand, daß sich die aus dem Beurteilungsverfahren ergebenden pädagogischen Konsequenzen nicht realisieren lassen. Ihre Realisierung scheitert zum einen an den bestehenden institutionalisierten Ausbildungsalternativen, deren Anzahl in der Regel nicht höher als drei ist (z.B. die weiterführenden Ausbildungszweige nach der Grundschule oder die Differenzierungskurse innerhalb der weiterführenden Schulen). Sie scheitert zum anderen an den gesellschaftlichen, organisatorischen und finanziellen Barrieren, die die Entwicklung eines derartig komplexen Ausbildungssystems als Utopie erscheinen lassen. .

Im Hinblick auf die gegebenen Bedingungen des bestehenden Schulsystems mag die Beschäftigung mit dem ersten Rekonstruktionsmodell praktizierter Diagnostik in der Schule nachträglich als Spielerei erscheinen. Es gilt jedoch festzuhalten, daß die utopischen Konsequenzen für die Verkomplizierung des Ausbildungssystems aus den verfahrensimmanenten Prinzipien und Annahmen des Beurteilungsprozesses resultieren, der in den genannten curricularen Vorschriften - den vier Kriterien - seine allgemeine und damit verbindliche Grundlage hat. Ihre Unhaltbarkeit für die diagnostische Rekonstruktion von dispositionellen Eigenschaften von Schülern wurde bereits diskutiert. Zugleich kann man aber feststellen, daß das Vorgehen in der Beurteilung von Aufsätzen, wenn auch mit voneinander abweichenden Konkretionen, real existiert, und daß daraus, wegen der gegebenen, dazu widersprüchlichen Bedingungen des Ausbildungssystems nicht die erforderlichen Konsequenzen gezogen werden.

Es ist daher zu fragen, wie ein Lehrer, der bei der Aufsatzbeurteilung der Verfahrensform des ersten Rekonstruktionsmodells folgt, die Fülle der erhaltenen Daten im Sinne selektionsfunktionaler, diagnostischer Urteilsbildung zu synthetisieren vermag. Dieser Prozeß soll wiederum modellhaft rekonstruiert werden und stellt, weil der auf dem ersten Modell aufbaut, eine Erweiterung dessen dar. In der Betrachtung des zweiten Modells wird sich ergeben, daß die eingangs gemachten Voraussetzungen

preiszugeben sind und die Preisgabe selbst durch die Bedingungen der Selektionsentscheidung erzwungen wird.

Rekonstruktionsmodell II

Zur Darstellung des zweiten Rekonstruktionsmodells beziehe ich mich auf die Tabelle 1 und die für den Umgang mit den Symbolisierungen 1 und 0 geltenden Einschränkungen. Es wurde dort gesagt, daß aufgrund der für das erste Rekonstruktionsmodell geltenden Voraussetzungen, die Verknüpfung der symbolisierenden Ziffern 1 und 0 durch mathematische Operationen innerhalb eines Musters ebenso gegenindiziert ist wie ein darauf aufbauender Vergleich zwischen verschiedenen Mustern. Diese Einschränkung wird durch das Vorgehen in der Datenreduktion innerhalb des zweiten Rekonstruktionsmodells de facto für irrelevant erklärt.

Unter der Voraussetzung, daß die Aufsatzbeurteilung selektionsfunktional sein soll und daß deshalb nur soviele Schülergruppen durch Urteil unterschieden werden müssen wie Selektionsalternativen bzw. institutionalisierte Ausbildungsalternativen vorhanden sind, muß man durch Addition aller Ziffern innerhalb eines Musters die Häufigkeit der Merkmalsausprägungen "Vorhanden" charakterisieren. Man ist dann in der Lage, mit der dadurch erhaltenen Scheinkombination der vier Teilkriterien zu einem Gesamtkriterium so zu operieren, als läge ein komparativer Begriff vor, mit der sich die Aufsätze in gleiche und verschiedene einteilen lassen und die Verschiedenheit selbst durch eine Vorgängerrelation spezifiziert ist, d.h., Aufsätze mit größerer Häufigkeit der Merkmalsausprägung "Vorhanden" sind scheinbar besser als jene mit geringerer Häufigkeit.¹⁰⁾

10) Genaugenommen wird bei der Addition der Ziffern 1 und 0 so vorgegangen, als läge ein empirisch definierter, metrischer Begriff als Interpretation des Gesamtkriteriums vor, bei dem der Unterschied zwischen den Häufigkeiten nicht nur durch eine Vorgängerrelation spezifiziert ist, sondern durch eine metrische Abstandsrelation. Sogar diese Präzisionsstufe wird der schulischen Diagnostik in der Berechnung von Notendurchschnittswerten vorgetäuscht.

Wie aus der Tabelle 1 in der Spalte "Häufigkeit der Merkmalsausprägung 'vorhanden'" zu ersehen ist, sind dort die Muster bereits unter diesem Gesichtspunkt geordnet. Für den Aufsatz, der Muster 1 entspricht, konnte viermal die Merkmalsausprägung "vorhanden" konstatiert werden, bei Muster 16 in keinem Fall und bei den dazwischenliegenden Mustern entsprechend der in der letzten Spalte aufgeführten Zahl. Wie irrelevant für dieses Vorgehen eine inhaltlich begründete Analyse des Konstitutionsprozesses des Aufsatzes ist, zeigt sich daran, daß man die Zahlenreihe in der letzten Spalte durch Transformation mit dem Summanden 1 in die früher gültige Notenskala von 1 bis 5 umwandeln kann. Letzteres läßt sich auch durch eine veränderte Ziffernsymbolisierung der Merkmalsausprägungen "vorhanden" bzw. "nicht vorhanden" erreichen, indem man statt der Ziffern 1 und 0 die Zahlen 1,25 und 0,25 verwendet.

Beide Verfahrensweisen bringen auf einer sehr anschaulichen Ebene die reine **Darstellungsfunktion** der Symbolisierung zum Ausdruck und das Fehlen jeglicher Erkenntnisfunktion. Mit der symbolischen Darstellung ist keinerlei Erkenntniszuwachs verbunden, weil sie selbst nur implizite, formale Aspekte des diagnostischen Prozesses verallgemeinern kann. Sie verleiht aber gerade dadurch dem gesamten Vorgehen einen scheinbar höheren Grad von Plausibilität, zumal die vollzogene Datenreduktion für die Aufteilung der Schüler auf institutionalisierte Ausbildungsalternativen einfacher ist als eine sensibler differenzierende Methode.

Zugleich liefert die symbolische Formalisierung einen weiteren ausbildungsinstitutionellen Vorteil, der sich ebenso wie die Transformation der Häufigkeiten der Merkmalsausprägungen in Noten als formalmethodische Legitimation selektionsorientierter Diagnostik in der Schule erweist. Es läßt sich mathematisch zeigen, daß die Anzahl gleicher Häufigkeiten von Merkmalsausprägungen als Binomialkoeffizient durch die Anzahl der Kombinationen von vier Elementen zur k-ten Klasse dargestellt werden kann (vgl. RASCH, 1970). M.a.W.: Die Anzahl gleicher Häu-

figkeiten folgt einer theoretischen Verteilung (Binomialverteilung), die - wie auch die Normalverteilung - im Rahmen pragmatisch tollerierbarer Abweichungen von der konventionell realen Verteilung der Zensuren einer erwünschten Quotierung der Zulassung von Schülern in weiterführende Schulen eine zusätzliche Stütze liefert. Nicht selten wird daher die nachträglich festgestellte Ähnlichkeit zwischen der realiter erfolgten Zulassung bestimmter Quoten von Schülern in weiterführende Schulen und der Prozentsätze, die sich aus der theoretischen Verteilung ergeben, als Bestätigung der Hypothese aufgefaßt, Merkmalsausprägungen würden einer bestimmten Form einer theoretischen Verteilung folgen und das Ganze als ein empirischer Sachverhalt betrachtet. Objektiv ist die Ähnlichkeit dagegen nur eine Bestätigung des Sachverhaltes, daß sich der Klassifikationsprozeß in Bezug auf das ausgewählte Beurteilungsmerkmal nach den modellimmanenten Bedingungen der theoretischen Verteilung orientiert.

Am vorliegenden Beispiel der symbolisierten Aufsatzbeurteilung sind für die Häufigkeitsanzahl 4 theoretisch ca. 6,3% von Aufsätzen zu erwarten, die dieses Muster zeigen, für die Häufigkeitsanzahl 3 sind es 25%, für 2 ca. 38,4%, für 1 wiederum 25% und für 0 wie beim ersten Muster ca. 6,3%. Bei Transformation der Häufigkeitsanzahl mit dem Summanden 1 und Umkehrung der Reihenfolge ergibt sich eine Notenskala mit der Spannweite von 1 bis 5 und einer entsprechenden prozentualen Verteilung der Noten: die Note 1 und 5 ist in jeweils 6,3% der Fälle zu erwarten, die Note 2 und 4 mit 25% und schließlich die 3 mit 38,4%. Will man der gegenwärtig institutionalisierten Gewohnheit in der Zensurengebung noch weiter entgegenkommen und die Noten von 1 bis 6 vergeben, dann genügt es, die Anzahl der Kriterien um einen Klassifikationsgesichtspunkt zu erweitern, z.B. durch eine Rechtschreibkategorie oder durch eine Kategorie für grammatikalische Richtigkeit der Sätze. Man erhält dann insgesamt $2^5 = 32$ verschiedene Beurteilungsmuster, deren rechnerische Synthese die folgende Häufigkeitsverteilung für die

Zensuren erbringt: Die Note 1 und 6 wird in jeweils 3,13% der Fälle erwartet, die Note 2 und 5 tritt mit einer Wahrscheinlichkeit von 15,6% auf und die Note 3 und 4 mit 31,3%.

Es sei nochmals ausdrücklich hervorgehoben, daß es sich bei den angegebenen Prozentsätzen um erwartete, theoretische Quantitäten handelt, die die Wahrscheinlichkeit ausdrücken, mit der jene Beurteilungsmuster auftreten können, die sich lediglich durch die Häufigkeit der Merkmalsausprägung "vorhanden" unterscheiden und die auf der Beurteilung des Aufsatzes unter den verschiedenen dichotomisierten Klassifikationsgesichtspunkten aufbauen. Realiter weicht die Verteilung solcher Beurteilungsmuster auch dann von der theoretisch erwarteten ab, wenn sich die Beurteilung von Aufsätzen in der beschriebenen Weise vollzieht. Dafür sind eine Reihe von Gründen verantwortlich zu machen, die zum einen in der Festlegung der Klassifikationsgesichtspunkte bzw. ihrer Begriffsstruktur zu suchen sind. Zum anderen beruht die Abweichung auf der Gesamtheit der empirischen Bedingungen, unter denen Aufsätze im organisierten Lernen geschrieben werden und der diagnostischen Rekonstruktion von dispositionellen Eigenschaften dienen sollen, also z.B. Spezifika der untersuchten Schülerpopulation, die lehrinhaltliche Vermittlung der Richtlinien zur Abfassung eines Erlebnisaufsatzes, die Art des kommunikativen Bezuges der Schüler in der Klasse untereinander und das davon abhängige Modellernen, die zeitliche Konsistenz der Interpretation der Klassifikationsgesichtspunkte durch den beurteilenden Lehrer, die implizite Beeinflussung dieser Interpretation durch Vorurteile gegenüber Schülern usw.

Obwohl die Kenntnis der Wirksamkeit solcher Faktoren der wissenschaftlichen Auseinandersetzung mit der Problematik der Schülerbeurteilung genügend Anstoß für die kritischen Untersuchungen des Lehrerurteils geliefert hat und bei den Lehrern selbst hinreichende, subjektive Unzufriedenheit mit der konventionellen Zensurenggebung provozierte, übt die Möglichkeit der Anwendung des formalmethodischen Modells der Rekonstruktion dispositioneller Eigenschaften eine fast irrationale Faszination

auf die Urteilenden aus. Offensichtlich kompensiert die Untrüglichkeit der formalen Synthese vorliegender symbolisierter Daten die Unreflektiertheit der Datengengese, in der die zentralen Schwierigkeiten verborgen liegen. So hat sich im Zuge der Verwissenschaftlichung der Diagnostik in der Schule bei vielen Urteilern, seien dies die Lehrer selbst oder Wissenschaftler, die sich um ein objektiviertes Beurteilungsverfahren bemühen, die Tendenz entwickelt, eine durch Untersuchung erhaltene, aber von einer theoretischen verschiedene Verteilung der Merkmalsausprägungen nachträglich der theoretischen anzupassen oder durch Veränderung des Beurteilungsprozesses zu korrigieren. Als theoretische Verteilung fungiert dabei meistens die Normalverteilung, weil sie in standardisierter Form und durch ihre konventionalisierte Verwendung arbeitsökonomischer einzusetzen ist.

Die Empfehlung von wissenschaftlicher Seite, den diagnostischen Prozeß von Anfang an so zu gestalten, daß die resultierende Verteilung der Merkmalsausprägungen der Normalverteilung nahekommt, begleitet fast jede Einführung in die Konstruktion von "objektiven" Verfahren zu Schülerbeurteilung (LANGFELDT, 1974; HORN, 1974; MAGNUSSON, 1969); ganz zu schweigen von den Hinweisen auf die Notwendigkeit, sie vorauszusetzen, um die methodische Brauchbarkeit von Tests sicherzustellen. Darüberhinaus bleibt jedoch die sich darin abzeichnende Verselbständigung methodischer Prinzipien und Richtlinien unverständlich, wenn man ihre legitimatorische Funktion für die Durchsetzung schulischer Selektion nicht zur Kenntnis nimmt. "In Schweden, wo die Fragen der relativen (am Vergleich mit Mitschülern orientierten) und der absoluten (am Vergleich mit Leistungsstandards orientierten) Zensurengebung wohl am intensivsten diskutiert wurde, war die Anlehnung an die Normalverteilung zeitweilig sogar Vorschrift" (INGENKAMP, 1975, S. 42). Auch in Großbritannien ist die "statistische Angleichung" eine Methode zur Standardisierung selektionsbezogener Schülerbeurteilung (LALL, 1974). Daneben bewirken aber auch empfehlungsähnliche, kom-

mentierte Darstellungen der Normalverteilung in Arbeitsunterlagen für Lehrer (vgl. FEIGEL & KEITEL, 1967), daß diese ihre Zensurenvergabe an theoretischen Verteilungen orientieren.

Zusammenfassend betrachtet, rekonstruieren beide Modelle die Umkehrung der Methode auf der Ebene der konkreten Beurteiler-tätigkeit des einzelnen Lehrers; d.h.: Die begriffliche Klassifikation von vergegenständlichten Verhaltensweisen von Schülern, deren Beurteilung, wird unter formale Strukturen subsumiert, wohingegen die Entwicklung eines formalen Modells für ein allgemeines Vorgehen, das auf der inhaltlichen Analyse aufbaut, unterbleibt. An dieser Stelle ist zur Erörterung des Vorschlages von SCHMIDT zurückzukehren.

Vor dem Hintergrund seiner Forderung, daß für die diagnostische Rekonstruktion dispositioneller Eigenschaften die Untersuchung der Wechselwirkung zwischen dem Interaktionssystem des untersuchten Individuums und dem des Forschers bzw. Diagnostikers konstitutiv sein muß, kann die im zweiten Rekonstruktionsmodell stattfindende Umkehrung der Methode zunächst als ein Vorgehen charakterisiert werden, das von der Analyse des Interaktionssystem S_I völlig absieht. Die Beurteilung der Aufsätze beschränkt sich auf die Anlegung von Klassifikationsgesichtspunkten, die sich durch drei Bestimmungen kennzeichnen lassen:

1. Über die empirische Konstitution der Aufsatzbestandteile, mit denen die Klassifikationsgesichtspunkte im Prozeß der Beurteilung interpretiert werden, ist soviel wie nichts bekannt. Die Untersuchungen mit der Methode des "lauten Denkens" im Bereich der kognitiven Psychologie (NEWELL, SHAW & SIMON, 1967) sind zwar erste Ansätze zur Erforschung des Interaktionssystems S_I , sie geraten aber in dem Maße zu einem formalistischen Vorgehen, wie es nicht gelingt, in die Kategorisierung des "laut Gedachten" das untersuchte Individuum miteinzubeziehen. Sehr skeptisch muß in dieser Hinsicht auf der Versuch stimmen, den Konstitutionsprozeß der Verhaltensresultate computertechnisch zu simulieren, da hier von vornherin die Gefahr einer impli-

ziten Orientierung der Klassifikation des "Laut-Gedachten" nach den handhabbaren, technischen Vorgängen besteht. Insgesamt heißt dies, daß die Klassifikationsgesichtspunkte für die Beurteilung der Verhaltensresultate - im vorliegenden Fall der Aufsatz - ausschließlich im Interaktionssystem S_F entwickelt wurden.

2. Als Folge davon muß zum Zweck der Anwendbarkeit der Klassifikationsgesichtspunkte für deren intersubjektiv konsistenten Gebrauch innerhalb des Interaktionssystems S_F gesorgt werden, d.h. für eine eindeutige Interpretation. Erweist sich die Unbezogenheit der Beurteilungskategorien auf das Interaktionssystem des untersuchten Individuums als nicht zur Kenntnis genommenes Hindernis für die Gewinnung einer vermittelbar eindeutigen Interpretation, wenn diese der diagnostischen Rekonstruktion *dispositioneller* Eigenschaften des Individuums dienen soll, so gelingt die Konsensusbildung auch nicht im Hinblick auf den institutionell Verwendungszusammenhang der Urteile. Hier können - wie bereits oben ausgeführt - durch Konsensustechniken, Entscheidungsmodelle und Versuche der empiristischen Rekonstruktion des Bedeutungsgehaltes von Beurteilungskategorien nur zentrale Tendenzen der Interpretationsrichtung beschrieben werden. Jenseits dessen haben die Versuche, durch bildungsökonomische Ansätze für curriculare Entscheidungen Grundlagen für Qualifikationsanforderungen und damit letztendendes für Beurteilungskategorien zu schaffen, ohnehin versagt, den gesellschaftlichen Bedeutungsaspekt intersubjektiv zu vermitteln (vgl. OFFE, 1975).

Um es am Aufsatzbeispiel zu verdeutlichen: Es dürfte kaum möglich sein, einen begründeten Konsensus darüber herzustellen, ob und in welcher Weise bzw. in Bezug auf welche Anforderungen die Erfüllung der Bedingung "der Schüler muß die einzelnen Erlebnisschritte klar herausarbeiten können" für den Erfolg in einem der drei weiterführenden Schulzweige notwendig ist. Es dürfte auch nicht gelingen, die Erfüllung der Bedingung in einen rational rekonstruierbaren Zusammenhang mit bestimmten Qua-

lifikationsanforderungen gewisser Beschäftigungsbereiche zu stellen.

Generell kann man behaupten, daß der ausschließlich im Interaktionssystem S_F ablaufende Prozeß der Konsensusbildung sicher immer wieder an jenem berühmten "pragmatischen" Einwand gestöppt wird, der sich nachträglich nicht mehr spezifizieren läßt.

MEYER benennt einen Teilaspekt dieses Pragmatismus, wenn er nach der prinzipiellen Kritik an der verhaltenstheoretisch orientierten Normenanalyse im Rahmen der Curriculumentwicklung feststellt: "Eine zweite Einschränkung der Brauchbarkeit der Strategien ist nicht systematischer Natur, aber von forschungspraktischer Bedeutsamkeit. Es ist die Frage, ob die benötigten Informationen und die erhofften Erklärungen über Beurteilungsprozesse mit einem vertretbaren **A u s m a ß a n K o s t e n u n d F o r s c h u n g s a u f w a n d** und in der angemessenen zeitlichen Frist zu erwarten sind" (ebd., S. 183).

Was immer sich hinter dem pragmatischen Einwand noch verbergen mag, unter der Perspektive der versuchten Konsensusbildung scheitert die intersubjektiv konsistente Interpretation von Klassifikationsgesichtspunkten zur Beurteilung prinzipiell an der Beschränkung des Untersuchungsansatzes auf die Vorgänge im Interaktionssystem S_F . Es ist sogar so, daß die dadurch bedingte Unmöglichkeit, den empirischen Bedeutungsgehalt als objektiv reale Basis einer intersubjektiven Verständigung anzugeben, im Gegensatz zur intersubjektiven Verständigungsmöglichkeit in den exakten Naturwissenschaften, die **o r g a n i s i e r t e** Konsensusbildung erforderlich macht. Diese kann deshalb erneut zum Gegenstand wissenschaftlicher Auseinandersetzungen werden. Für den Ablauf und das Ergebnis des organisierten Vorgehens wie etwa in der Richtlinienvergabe, Curriculumplanung, Lernzieloperationalisierung und Leistungsmessung ist daher eine **F o r m** zu finden, mit der das Vorgehen selbst vermittelt werden kann.

Letzteres leitet zur dritten Bestimmung über und führt zugleich auf das Syntheseverfahren im zweiten Rekonstruktionsmodell zurück.

3. Es wurde vorher schon darauf hingewiesen, daß die Formalisierung eines Beurteilungsprozesses zwar keinerlei Erkenntniszuwachs erbringt, jedoch dem Resultat der Formalisierung trotz aller Abstraktheit gegenüber dem konkreten, durchgeführten Beurteilungsprozeß eine gewisse, scheinbar anschaulich gegebene Plausibilität anhaftet. Ich werde auf dieses eigentümliche Phänomen bei der Auseinandersetzung mit der funktionalen Bedeutung von methodischen Paradigmen zurückkommen.

Im Augenblick geht es um die Klärung der Frage, ob die symbolisierende Darstellung des Beurteilungsprozesses in der Art des zweiten Rekonstruktionsmodelles als Substitutionsmethode für die erfolglose Konsensusbildung in Bezug auf die eindeutige Interpretation der Klassifikationsgesichtspunkte fungiert und ob durch die damit erreichte Formalisierung der Interpretation ein erneuter Beleg für die Vernachlässigung des Interaktionssystems S_I geliefert wird. Die Klärung der Frage ist deshalb bedeutsam, weil mit der Interpretation der Anspruch auf eine empirische Deutung der Beurteilungskategorien erhoben wird und der empirische Bedeutungsaspekt eine im Schüler existierende Disposition repräsentieren soll. Solange man diesen Anspruch nicht erhebt, unterscheidet sich die Beurteilung eines Aufsatzes nicht von der exegetischen Betrachtung eines Textes.

Die Exemplifizierung des zweiten Rekonstruktionsmodelles zeigte, daß die scheinquantifizierende Kombination der verschiedenen Aufsatzkriterien innerhalb eines Beurteilungsmusters und der scheinquantifizierende Vergleich zwischen verschiedenen Beurteilungsmustern im Widerspruch zur Struktur der Kriterien steht. Die Beurteilungskriterien waren als klassifikatorische Begriffe eingeführt und als voneinander unabhängig betrachtet worden. Demgegenüber steht die ausbildungsinstitutionell durchgesetzte Notwendigkeit, Urteile über Aufsätze abzugeben, die einen Vergleich zwischen den Aufsätzen und damit zwischen den

Beurteilungsmustern zulassen, d.h., mit deren Hilfe man die Aufsätze graduell abstufen kann. Da die institutionellen Interessen - als durchgesetzte - Priorität für die Urteilsbildung haben, ist der Beurteilungsprozeß von Anfang an in den Verwendungskontext selektiver Entscheidungen über die Zulassung von Schülern in weiterführende Ausbildungsalternativen integriert. Das Formalisierungsverfahren im zweiten Rekonstruktionsmodell realisiert die Funktionalisierung der Urteilsbildung für diesen Verwendungskontext durch

1. die potentielle Kalkulierbarkeit der Prozentsätze mehr oder weniger brauchbarer Prüfungsergebnisse als Grundlage für die Selektionsentscheidung,
2. die methodische Synthese symbolisierter Beurteilungen zu numerischen Urteilen und
3. die Auswahl der Menge und Art von Klassifikationsgesichtspunkten, die den immanenten Bedingungen des Formalisierungsverfahrens am besten entsprechen.

Durch die rechnerische Synthese der numerisch symbolisierten Merkmalsausprägungen werden dabei letztere so behandelt, als wären sie meßbare Eigenschaften des Aufsatzes, die deshalb zu kombinieren sind, weil sie angeblich eine einheitliche Dimension des Aufsatzes operationalisieren. Die konstitutiven Voraussetzungen für die Klassifikation des Aufsatzes sind somit preisgegeben und die Abstufbarkeit der verschiedenen Aufsätze scheint die Interpretation der numerischen Beurteilungsergebnisse als Ausprägungsgrade einer dispositionellen Eigenschaft der Schüler plausibel zu machen. Empiristisch wissenschaftstheoretisch gesehen wurde auf diesem Weg die Einführung eines metrischen Prädiaktes in die Wissenschaftssprache durch die Festlegung einer eigentlich erst danach entwickelbaren Skalierungsvorschrift stillschweigend ersetzt.

Diese "Umkehrung der Methode" ist wiederum eine Konsequenz der Beschränkung der Rekonstruktion dispositioneller Eigenschaften aus den Vorgängen im Interaktionssystem S_F . Die Besonderheit der Beschränkung stellt sich hier als Formalisierung der Interpretation der Klassifikationsgesichtspunkte dar. Scheinbar von

aller Subjektivität in der Deutung der Beurteilungskategorien befreit, ermöglicht sie eine Standardisierung der Beurteilungsergebnisse - die Darstellung der Aufsatzbeurteilung in Form einer Ziffer. Sieht man von der realen Konstitution solcher Ziffern ab und faßt nur ihre mathematischen Eigenschaften ins Auge, z.B. die für sie festgelegten Ordnungsrelationen bzw. metrischen Strukturen, dann eignen sie sich arbeitsökonomisch vorzüglich für eine rationelle Organisation des Selektionsverfahrens, wenn ihnen vorher die einzelnen Schüler zugeordnet wurden. Unter der Perspektive der rationalen Organisation ist es dann auch nicht mehr notwendig, ihren semantischen Gehalt zu hinterfragen; denn das gesamte Verfahren bewährt sich umso mehr, je öfters es durchgeführt wird.

Die Bewährungsmöglichkeit ist aber durch einen Teilfaktor im Interaktionssystem S_F vorgegeben. Im Verwendungskontext "selektives Ausbildungssystem" liegt der konventionellen Zensurengebung und auch den Versuchen, diese durch objektivierte diagnostische Verfahren zu ersetzen, eine dafür wichtige Voraussetzung zugrunde: Die zur Verfügung stehenden weiterführenden Ausbildungsalternativen bzw. bestimmte Formen der Unterrichtsdifferenzierung stellen graduell abstufbare Qualifikationsanforderungen an die aufzunehmenden Schüler und "erfahrungsgemäß" entspricht nur eine zunehmend geringere Menge von Schülern wachsenden Anforderungen. So deutlich sich diese Annahme in der konventionellen Organisation des Lernens und der institutionalisierten Form des Ausbildungssystems immer wieder zu bestätigen scheint und unter den bestehenden Bedingungen vollkommen richtig ist, so wenig ist sie allgemein aufrechterhaltbar, wenn man sie als Resultat der geschichtlichen Entwicklung des Bildungswesens begreift.

Bereits eine oberflächliche Analyse der historischen Entstehung des Schulsystems beweist die Tatsache, daß die organisierte Ausbildung im allgemeinen wie auch die privilegierte Ausbildung und Bildung im besondern zuerst für eine privilegierte Schicht von wenigen geschaffen wurde und erst im Zuge der zunehmenden

Vergesellschaftung der Qualifikationsbedürfnisse eine Verallgemeinerung erfährt (vgl. HARTMANN, NYSSSEN & WALDEYER, 1974). Aus historischer Sicht ist also die Formulierbarkeit der Annahme eine Folge der objektiven Entwicklung des Ausbildungswesens.

Wissenschaftstheoretisch ist die Annahme insofern problematisch, als zwar die objektiven Bedingungen der historischen Entwicklung in sie eingehen, aber dort nicht explizit als Bedingung für ihre Aufrechterhaltbarkeit in Erscheinung treten. Das heißt: Die Annahme, daß nur zunehmend wenige Schüler in Ausbildungsalternativen mit wachsenden Anforderungen Erfolg haben werden, ist nur dann haltbar, wenn die Tatsache unbenannt bleibt, daß in diesen Ausbildungsalternativen historisch tradierte Unterrichtsstandards durchgesetzt sind, die sich erst allmählich durch den Unterricht mit bestimmten privilegierten und wenigen Schülern entfaltet haben. Dieser Gedanke wird in der Problematisierung des pädagogischen Gesetzesbegriffes durch ITELSON (1967) systematisch präzisiert. Danach kann man die Annahme wissenschaftstheoretisch als eine Hypothese auffassen, in der eine nichtexplizierte Existenzbedingung von der Art enthalten ist, wonach der unterschiedliche Erfolg von nur wenigen durch einen spezifischen Unterrichtsmodus bedingt ist. Wird diese Existenzbedingung nicht expliziert, dann ist die Hypothese immer richtig, weder zu belegen noch zu entkräften und damit tautologisch. Gleichwohl ist diese Annahme die Legitimation für die unbegründbare Forderung, einen (schein-)quantifizierenden Beurteilungsmodus für die Klassifikation von vergenständlichten Schülerverhaltensweisen zu schaffen, deren Einlösung in die diskutierte Widersprüche führt.

Hier ist nun zu erkennen, daß die Formalisierung der Beurteilung im zweiten Rekonstruktionsmodell auf tradierte Annahmen im Interaktionssystem S_F zurückzuführen ist; auf Annahmen, die die Voreinstellung von Diagnostikern bzw. Forschern prägen und deren objektiver Hintergrund nur in den existierenden, gesellschaftlichen Strukturen des institutionalisierten Ausbildungssystems in seiner historischen Gewordenheit zu suchen ist. Aus

dem Interaktionssystem S_I kann die Voreinstellung überhaupt nicht bezogen worden sein, weil zum einen Analysen darüber bis heute noch nicht vorliegen und zum anderen, weil die Forderung nach quantifizierender Beurteilung eine transformierte Kennzeichnung i n s t i t u t i o n a l i s i e r t e r Ausbildungsverhältnisse darstellt.

Ihre versuchte Erfüllung hat allerdings den entscheidenden "Vorteil", daß sich formalisierte Beurteilungsergebnisse, eben, weil sie scheinbar von aller Subjektivität befreit sind, auch dann transformieren bzw. intersubjektiv vermitteln und weitergeben lassen, wenn konkrete Personen die Informationsträger sind. Da die Form der Beurteilung - im Bereich der praktizierten Diagnostik in der Schule ist das die optimale Unterscheidung von Schülern zum Zweck der Verteilung von Bildungschancen - nur der Struktur des institutionalisierten Ausbildungswesens entspringt und nur in diesem Rahmen ihre begrenzte Entwicklungsmöglichkeit erfährt, ist sie die Grundbedingung für das Entstehen der aufweisbaren Widersprüche. Sie repräsentiert die zentrale Ursache für die Beschränkung der diagnostischen Rekonstruktion von dispositionellen Eigenschaften auf die Rationalisierung der Vorgänge im Interaktionssystem S_F und es wird deshalb später zu fragen sein, wie eine methodisch objektivierte Beurteilung dem Dilemma ausweichen will.

2.4 Der Verlust des Anwendungsbereiches pädagogischer Diagnostik und die gemeinsame Verständigungsbasis verschiedener Beurteiler

Nach der Darstellung des Modells für die Rekonstruktion dispositioneller Eigenschaften als typisierte Form der praktizierten Diagnostik in der Schule anhand der Zensurengebung und dem Aufweis der dabei auftretenden Schwierigkeiten ist es notwendig, wieder zur Frage nach der Eigenart diagnostischer Dispositionsprädikate zurückzukehren. Durch ihre systematische Untersuchung wird die theoretische Voraussetzung für die kritische Beurteilung der Versuche geschaffen, die praktizierte Diagnostik in der Schule zu objektivieren.

Im Anschluß an die Erörterung physikalischer Dispositionsprädikate und ihrer Kennzeichnung als Rekonstrukte - was auch für die diagnostischen Dispositionsprädikate Gültigkeit hat - wurde die Frage gestellt, ob Dispositionsprädikate als Ergebnis der diagnostischen Rekonstruktion von Schülereigenschaften die gleiche Struktur und Konstitution wie die physikalischen haben. Ein expliziter Vergleich beider Dispositionsbegriffe hat bis jetzt noch nicht stattgefunden, er wurde aber durch die Vorstellung des Forschungsansatzes von SCHMIDT sowie durch die Beurteilung der beiden Rekonstruktionsmodelle mit diesem Forschungsansatz vorbereitet.

Allgemein soll mit Dispositionsbegriffen die invariante Tendenz von Gegenständen sprachlich repräsentiert werden, sich unter angebbaren Bedingungen in angebbare Zustände zu transformieren. Beide Dispositionsbegriffe - die diagnostischen und die physikalischen - lassen sich durch jene Operationen unterscheiden, die der Forscher sprachlich zu rekonstruieren versucht:

1. Im Fall der physikalischen Dispositionsbegriffe rekonstruiert der Forscher sprachlich Operationen, die sich am Gegenstand vollziehen und ohne sein Zutun in naturgesetzlicher Form ablaufen (z.B. der Zucker löst sich im Wasser auf) und Opera-

tionen, die er selbst ausführt, damit erstere ablaufen können (z.B. er baut die Versuchsanordnung auf und wirft den Zucker in das Wasser). Ist der gesamte Prozeß grundsätzlich reproduzierbar, d.h., besteht eine feststellbare Kontingenz zwischen beiden Operationenkategorien, dann kann dies als objektive Grundlage für die Einführung eines Dispositionsbegriffes in die Wissenschaftssprache aufgefaßt werden.

2. Im Fall des diagnostischen Dispositionsbegriffes rekonstruiert der Forscher zwar auch sprachlich Operationen, die er selbst durchführt wie z.B. die Darbietung einer Aufgabe ("Schreibe einen Aufsatz mit diesem oder jenem Thema") oder die Konstruktion und Darbietung mehrerer Testitems in Form eines Schulleistungstests. Er ist aber nicht in der Lage, Operationen zu rekonstruieren, die sich in Entsprechung zum physikalischen Geschehen quasi-naturgesetzlich am Gegenstand vollziehen; weder am Gegenstand "Schüler", da hier die Aufgabenlösungsoperationen nicht anschaulich gegeben bzw. erlebnismäßig aufweisbar sind, noch am Gegenstand "Aufgabenlösungsergebnis" oder "Verhaltensresultat", da hier die Operationen bereits zu einem statischen Produkt geronnen sind.

Außerdem steht er bei der Rekonstruktion von Dispositionen vor der Aufgabe, zwei Arten von Gegenstände in die Einführung des Dispositionsbegriffes miteinzubeziehen: den die Lösungsoperationen ausführenden Schüler und dessen Vergegenständlichung, das vom Schüler getrennte Lösungsergebnis. Schließlich ist ein zusätzlicher Unterschied in der Einführung eines diagnostischen Dispositionsprädiaktes gegenüber der Einführung eines physikalischen noch darin zu sehen, daß die Konfrontation des Schülers mit der vom Forscher bereitgestellten diagnostischen Situation keinesfalls einen eindeutigen und bekannten Veränderungsprozeß im Schüler auslöst, dessen Wiederholbarkeit die Voraussetzung für die Annahme einer dispositionellen Eigenschaft abgeben könnte. Für den Forscher liegt im günstigsten Fall eine Art von Erfahrung vor, daß frühere Darbietungen der gleichen diagnostischen Situation bei einer begrenzten Menge von Schülern zu ähnlichen Lösungsergebnissen führte.

Aus der Unbekanntheit der Lösungsoperationen, die durch die diagnostische Situation provoziert werden, und aus der Getrenntheit des Gegenstandes, der die Lösungsoperationen durchführt, von jenem Gegenstand, der das Ergebnis der Lösungsoperationen darstellt, ergeben sich eine Reihe von Einschränkungen für die *a n a l o g i s i e r e n d e* Übertragung der Einführung physikalischer Dispositionsbegriffe in die Wissenschaftssprache auf die der diagnostischen.

2.4.1 Die kompliziertere Struktur diagnostischer Dispositionsprädikate

In der diagnostischen Rekonstruktion von dispositionellen Eigenschaften geht es nicht mehr um das Problem, wie die Allgemeingültigkeit einer Aussage von anschaulich gegebenen und reproduzierbaren, und damit invarianten Gegenstandstransformationen, die nur in begrenztem Ausmaß wiederholt werden, nachzuweisen ist. Hier sind gleichzeitig zwei Fragen zu beantworten:

Ist das Phänomen, daß verschiedene vergegenständlichte Verhaltensweisen eines Schülers in einer diagnostischen Situation in gleicher Weise klassifiziert werden können, auch auf dispositionelle Eigenschaften als invariante, in der diagnostischen Situation wirksame Ursachenkomplexe zurückzuführen, oder nur auf das Klassifikationsvorgehen des Forschers ?

Ist die sprachlich rekonstruierte Invarianz für die diagnostische Situation spezifisch oder ist die Aussage "Wenn ich dem Schüler äquivalente Aufgaben vorlege, dann produziert er äquivalente Lösungen" allgemeingültig und das reproduzierte Lösungsverhalten eine prinzipiell invariante, dispositionelle Eigenschaft des Schülers ?

Die letzte Frage ist mit jener vergleichbar, die sich bei der Einführung von physikalischen Dispositionsbegriffen in die Wissenschaftssprache stellt. Die erste impliziert die inhalt-

liche Aufklärung der festgestellten Invarianz, die für die Beantwortung der zweiten vorausgesetzt werden muß. Bei der Einführung von physikalischen Dispositionsbegriffen kann man sie voraussetzen, weil der anschaulich gegebene Transformationsprozeß des Gegenstandes selbst die empirisch objektive Grundlage für die inhaltliche Bedeutung des Prädiaktes z.B. "Wasserlöslichkeit" liefert. Bei der Einführung von diagnostischen Dispositionsprädikaten läßt sich von dieser Voraussetzung nicht ausgehen, weil sich hier der Transformationsprozeß des Gegenstandes "Schüler" einer beobachtungssprachlichen Verallgemeinerung entzieht und deshalb die inhaltliche Bedeutung der zu rekonstruierenden, dispositionellen Eigenschaft des Schülers ausschließlich von der Klassifikation der diagnostischen Situation und der produzierten Lösungsergebnisse abhängt.

Wird dabei wie im Fall der selektionsorientierten Diagnostik der Klassifikation des Forschers bzw. Diagnostikers Priorität eingeräumt, so müßte für die inhaltliche Aufklärung festgestellter, invarianter Beziehungen zwischen Prüfungsaufgaben und Lösungsergebnissen bei einem Schüler zuvor überprüft werden, ob der Schüler die Klassifikation in gleicher Weise vornimmt wie der Forscher. Es müßte also untersucht werden, ob die Forderung von SCHMIDT erfüllt ist oder nicht. Unterschwellig deutet sich die Problematik der Vermittelbarkeit der Klassifikationsgesichtspunkte zwischen den beiden Interaktionssystemen S_I und S_F sogar in der naiv-empiristischen Unterrichtsforschung an, wenn etwa die These aufgestellt wird, der Unterrichtserfolg des Lerners sei praktisch schon dadurch garantiert, daß der Lehrer den Schülern seine Unterrichtsziele erklärt (MAGER, 1971).

Solange die inhaltliche Aufklärung von festgestellten, invarianten Beziehungen zwischen diagnostischer Situation und klassifizierten Verhaltensresultaten in dieser spezifischen diagnostischen Situation nicht erfolgt, besteht überhaupt keine Möglichkeit, die Beziehung empirisch zu begründen und über die Situation hinaus zu generalisieren. Denn dann wird mit der

unternommenen Bestätigung der Hypothese "Wenn ich dem Schüler äquivalente Aufgaben vorlege, dann produziert er äquivalente Lösungen" die Antwort auf beide Fragen zu geben versucht und ein sprachlich repräsentiertes Phänomen mit seiner Allgemeingültigkeit gleichgesetzt. Dieses naiv-empiristische Vorgehen steht in eindeutigen Gegensatz zur geforderten Vorgehensweise in der Einführung von Dispositionsbegriffes durch Aussagen der genannten Art. Mit DICK (1974) kann man behaupten, daß hier ein empirisches Phänomen zugleich Ort der Anpassung und Bestätigung der Aussage ist; Ort der Anpassung, als daß konkrete, diagnostische Operieren sprachlich verallgemeinert wird, andererseits aber Bestätigung, als das gleiche diagnostische Operieren eine erfüllende Konkretion der Aussage darstellen soll.

Im wesentlichen hängt also die Nichtübertragbarkeit des Einführungsverfahrens bei physikalischen Dispositionsbegriffen auf diagnostische mit der Unbestimmtheit der Operationen zusammen, die der Schüler beim Lösen von Aufgaben ausführt. Begriffslogisch spezifiziert sich der Unterschied in der Struktur der beiden Dispositionsprädiakte: Das physikalische ist ein zweistelliges; es wird durch einen Wenn-Dann-Satz eingeführt. Dies gilt auch dann, wenn der einführende Wenn-Dann-Satz nicht als Operationalisierung (BRIDGMAN, 1951) aufgefaßt wird, sondern als Zuordnungsregel für die Verbindung zwischen dem Dispositionsbegriff und einer Beobachtungssprache (vgl. STEGMÜLLER, 1970, Band II, Studienausgabe Teil B und Teil C).

Das diagnostische Dispositionsprädikat ist dreistellig; aus diesem Grund ist der obige Satz zu vervollständigen und müßte lauten: "Wenn ich dem Schüler äquivalente Aufgaben vorlege, und er setzt sich damit in einer bestimmten Weise auseinander, dann produziert er äquivalente Lösungen". Erst die Kenntnis der "bestimmten Weise" liefert die Grundlage für die inhaltliche Interpretation der festgestellten, invarianten Beziehung. Ohne sie bleibt die Interpretation zirkulär und ermöglicht ihre subjektivistische Auffüllung.

2.4.2 Die fehlende Verfügung über eine Beobachtungssprache

Der nächste Einwand gegen die Übertragung des Einführungsschemas von Dispositionsbegriffen baut auf dem vorigen auf und schließt an die Diskussion empiristischer Versuche an, eine intersubjektiv konsistent gebrauchte Beurteilung der Verhaltensresultate zu finden. Ich gehe von der Voraussetzung aus, man würde sich innerhalb der empiristisch orientierten Forschung die Mühe machen, die diagnostische Rekonstruktion von dispositionellen Eigenschaften in Anlehnung an das formalisierte Interaktionsmodell von SCHMIDT durchzuführen. Dabei ist zu berücksichtigen, daß das SCHMIDTsche Modell lediglich formale Bedingungen für die Untersuchung des diagnostischen Prozesses und für das Untersuchungsergebnis benennt, nicht aber die Forschungsmethode selbst festlegt.¹¹⁾

-
- 11) Forschungsmethodisch wäre zur Einlösung seiner Forderung vor der empiristischen Rekonstruktion von dispositionellen Eigenschaften eine Untersuchung durchzuführen, die die Überprüfung der Kongruenz der Interpretation der gesamten diagnostischen Situation durch den Forscher und durch das untersuchte Individuum zum Ziel hat. Dieses Ziel ließe sich nicht dadurch erreichen, daß die Interpretationsweise des untersuchten Individuums sukzessiv durch feiner differenzierende Urteilskategorien des Forschers bzw. durch objektivierte Beurteilungsverfahren strukturiert würde. Der Untersuchungsweg müßte vielmehr seinen Anfang von dem Vermittlungsversuch vorläufiger, subjektiver Interpretationen der diagnostischen Situation in beiden Interaktionssystemen S_I und S_F nehmen, sich über den Versuch der Vermittlung gewonnener, gemeinsamer Kategorien an andere untersuchte Individuen und Forscher fortsetzen und könnte erst zum Schluß in der Formulierung von empirischen Hypothesen münden. In solchen empirischen Hypothesen wird ja letztlich nichts anderes behauptet, als daß sich die darin formulierte allgemeine Aussage wieder konkretisieren läßt. N.a.w.: Es wäre zu untersuchen, ob die subjektiv verschiedenen Bedeutungen der diagnostischen Situation für Forscher und untersuchtes Objekt zwischen beiden zu vermitteln sind und der Vermittlungsprozeß selbst zu rekonstruieren ist. Erst durch diese Untersuchung wäre die subjektive Sichtweise des einzelnen Forschers bzw. der Forschergemeinschaft zu konkretisieren (vgl. HOLZHAUSEN, 1973, S. 45-56).

Infolgedessen könnte ein empiristischer Wissenschaftler die Kongruenzbedingung, die für die Abbildung von Reizlage und Verhaltensresultat im Interaktionssystem S_I und S_F geltend gemacht wird, durch die Konstruktion eines komplexen Variablensystems einzulösen versuchen. Gemäß diesem System wäre die Strukturierung von Reizlage und Verhaltensresultat durch das untersuchte Individuum durch einen formalisierten Zusammenhang von verschiedenen, objektiviert erfaßten Merkmalen darzustellen. Zur Veranschaulichung des Gedankens benutze ich das Faktorenmodell für die "Struktur des Intellekts" von GUILFORD (1956).

Angenommen, der Schüler hätte eine Aufgabe zu lösen, die nach Auffassung des Forschers bzw. Diagnostikers im GUILFORDschen Modell folgendermaßen einzuordnen ist: Sie verlangt das Umgehen mit semantischen Denkinhalten, erfordert als Denkprodukte das Herstellen von Klassen und benötigt zu ihrer Lösung die Denkoperation "konvergente Produktion". Eine solcher Aufgabe könnte etwa heißen: "Was ist das Gemeinsame zwischen einem Führerschein und einem Fahrzeugschein?". Nehmen wir weiterhin an, die richtige Lösung der Aufgabe hieße "Zulassungspapiere" und die Aufgabe wäre eine der diagnostischen Bedingungen für die Untersuchung intelligenten Verhaltens.

Würde ein Schüler mehrere Aufgaben dieser Sorte lösen, dann käme ein aufgeklärter Empirist - jener, der nicht mehr den radikalen Verifikations- bzw. Falsifikationsstandpunkt vertritt - zu der Auffassung, daß der Schüler die Aufgaben deshalb richtig gelöst hat, weil er sie im Sinne ihrer Konzeption im Rahmen des GUILFORD-Modelles strukturiert hat; vorausgesetzt, das Überprüfungsverfahren könnte als partielle, empirische Deutung des unterstellten Intelligenzaspektes interpretiert werden. Bei einem anderen Schüler, der in der Lösung der Aufgaben erfolglos bleibt, müßte er annehmen, daß dieser die Aufgaben nicht im Sinne der Konzeption strukturiert hätte. Würde dieser Schüler nun bei der Lösung von Testaufgaben Erfolg haben, die konvergente Produktion und Herstellen von Klassen in Bezug auf symbolische Denkinhalte erfordern, dann wäre das Lösungsverhal-

ten des Schülers in einem ersten Schritt näher aufgeklärt und die Struktur der Disposition "intelligentes Verhalten" bei diesem Schüler genauer differenziert. Dies gilt wiederum nur unter der Voraussetzung, daß auch der zweite Test als partielle empirische Deutung des zweiten unterstellten Intelligenzasspektes aufzufassen ist.

Das Verfahren ließe sich auf diese Weise beliebig lange fortsetzen und würde zur Konzipierung eines komplizierten Variablen-systems führen, mit dem die dispositionelle Struktur der Auseinandersetzung verschiedener Schüler mit den diagnostischen Bedingungen abgebildet werden könnte. Auf die im Verfahren enthaltenen, ungelösten Schwierigkeiten bei der Auswahl und Begründung der Variablen und ihrer methodischen Synthese soll es an dieser Stelle nicht ankommen. Dies sind alles Probleme, die mit dem ersten Einwand zusammenhängen bzw. mit der Begründbarkeit des Skalenniveaus, das für die Erfassung der Merkmale angenommen wird. Wichtig ist vielmehr die Diskussion der zweimal genannten Voraussetzung, nach der die eingesetzten Prüfungsaufgaben als partielle, empirische Deutung von zu beurteilenden, dispositionellen Eigenschaften aufzufassen sind.

Damit die Voraussetzung gegeben ist, wird bei der Einführung von physikalischen Dispositionsbegriffen verlangt, daß die Aussagen, die zur Einführung des Dispositionsprädikates herangezogen werden, in einer intersubjektiv konsistent gebrauchten Beobachtungssprache formuliert sind. Wenngleich auch hier die Beantwortung der Frage, welche Kriterien eine intersubjektiv konsistent gebrauchte Beobachtungssprache erfüllen muß, noch offen ist (STEGMÜLLER, 1970, Band II, Studienausgabe Teil C), so bildet die erlebnismäßige Aufweisbarkeit der Operationen, die der Forscher am Gegenstand vollzieht und die am Gegenstand in naturgesetzlicher Form ablaufen, eine empirisch objektive Grundlage für die Gewinnung von intersubjektiver Konsistenz im Gebrauch einer beobachtungssprachlichen Verallgemeinerung des Geschehens - des interaktiven Geschehens zwischen Forscher und Gegenstand.

Bei der Einführung von diagnostischen Dispositionsbegriffen muß man aber von der Beschreibung der diagnostischen Anregungsbedingungen und der Beschreibung des Lösungsergebnisses ausgehen, also von der Beschreibung der Reizlage und der Verhaltensresultate als Indikatoren für die diagnostische Rekonstruktion dispositioneller Eigenschaften. Es ist daher zu fragen, ob man für beide Beschreibungen eine empirisch objektive Grundlage in dem Sinne voraussetzen kann, daß durch sie die beiden Arten von Operationen beobachtungssprachlich verallgemeinert werden kann - für die Beschreibung der diagnostischen Anregungsbedingungen jene Operationen, die der Forscher bzw. Diagnostiker zu ihrer Herstellung und Anwendung durchführt, für die Beschreibung der Lösungsergebnisse jene, die das untersuchte Individuum in der Auseinandersetzung mit den Anregungsbedingungen vollzieht.

Für die erste der beiden Beschreibungen kann man die Frage eindeutig beantworten. In den diagnostischen Anregungsbedingungen werden keine Operationen beschrieben, die die Konstruktion der Anregungsbedingungen betreffen, oder Operationen, die der Forscher am untersuchten Individuum auszuführen gedenkt. Es werden darin auch keine Operationen benannt, die im Gegensatz zur physikalischen Beobachtungssprache zu eindeutigen und bestimmten Transformationen des Gegenstandes führen.

Diagnostische Anregungsbedingungen wie "Schreibe einen Aufsatz mit diesem oder jenem Thema", "Löse die Rechenaufgabe von der folgenden Art" oder auch nur "Gegeben ist diese oder jene Aufgabe: was ist die richtige Lösung?" sind imperativische Aussagen, die den Schüler auffordern, sich damit in einer unterstellten, spezifischen Weise auseinanderzusetzen, wobei weder der Auseinandersetzungsprozeß noch sein Ergebnis im E i n z e l f a l l vorherzusagen ist. Sie setzen überdies voraus, daß der Schüler die Auseinandersetzung überhaupt intendiert. Sie verallgemeinern lediglich die implizite, aber nicht genannte Erfahrung, daß andere Schüler oder der betroffene selbst sich früher damit auseinandergesetzt und dabei unterschiedliche Lösungsergebnisse

tate erzielt haben. Nur in Bezug darauf bilden sie eine objektive Grundlage für ihren intersubjektiven Gebrauch innerhalb der Diagnostiker- bzw. Forschergemeinschaft.

Da über ihre empirische Bedeutung als notwendige Bedingung für die Auslösung eines b e k a n n t e n Transformationsprozesses des Gegenstandes - hier des untersuchten Individuums - nichts auszumachen ist und da genau umgekehrt, ihre inhaltliche Bedeutung aus den klassifizierten Lösungsergebnissen zu bestimmen versucht wird, verallgemeinern sie nur die Standards der Anregungsbedingungen, deren konkrete Wirkung auf das untersuchte Individuum unbekannt bleibt. Auch wenn sich der Vorgang, bei dem der Diagnostiker eine Prüfung vorbereitet und durchführt, in einer Beobachtungssprache ausdrücken läßt, so sind die damit beschriebenen Operationen für die Einführung eines Dispositionsbegriffes bedeutungslos; denn die Aktivierung von Dispositionen hängt von der Interpretation der diagnostischen Bedingungen durch das untersuchte Individuum ab, und die wird durch diese Art der beobachtungssprachlichen Beschreibung der Tätigkeiten des Diagnostikers nicht erfaßt.

Ein solches Problem stellt sich bei der Einführung von physikalischen Dispositionsbegriffen nicht, weil der untersuchte Gegenstand keine subjektive Interpretation der situativen Bedingungen vornimmt. Die Übertragung des Einführungsverfahrens auf diagnostische Dispositionsbegriffe scheitert daher schon deshalb, weil in die Festlegung einer intersubjektiv konsistent gebrauchten Beschreibung der diagnostischen Anregungsbedingungen auch deren Interpretation durch das untersuchte Individuum eingehen muß.

Solange dies fehlt, unterhalten sich Forscher nur über die formale Konstruktion und Durchführung von Prüfungsverfahren und über einen fiktiven, intendierten Anwendungsbereich dieser Verfahren. Daß dem so ist, und daß dadurch dem Problem auszuweichen versucht wird, die spezifische Bedeutungshaftigkeit der Anregungsbedingungen für das untersuchte Individuum zu berücksichtigen, belegen die Optimierungsbestrebungen im Bereich der

methodisch objektivierten Diagnostik. Die Forschergemeinschaft der empiristisch orientierten Diagnostiker bemüht sich um eine hinreichende Standardisierung des Testmaterials, der Testanweisung und der Durchführungsbestimmungen, um den Interpretationsspielraum des untersuchten Individuums sukzessiv einzuzengen und um dadurch einen verbesserten Konsens hinsichtlich der Beschreibung der Anregungsbedingungen zu erhalten. "Alle erwähnten Maßnahmen sollen zur Homogenisierung der Testsituation beitragen und der Sicherung der Vergleichbarkeit der Ergebnisse sowie der Erhöhung ihrer diagnostischen Bedeutsamkeit dienen" (VUKOVICH, 1964, S. 124).

Die Standardisierung des Testmaterials, der Testanweisung und Durchführungsbestimmungen - sie stellt den Vorgang dar, der häufig fälscherlicherweise als "Operationalisierung" bezeichnet wird - trägt allerdings nur mittelbar zur Erhöhung der diagnostischen Bedeutung der Testergebnisse bei. Je restriktiver die gesamte diagnostische Situation gestaltet wird, desto geringer ist der explizite Interpretationsspielraum des Diagnostikers für die Deutung der diskriminierenden Testergebnisse. Der Beitrag der Standardisierung läßt sich also nur negativ durch die Einschränkung der Deutung kennzeichnen und erbringt noch keine Positivierung der Interpretation der Testergebnisse. Mit der vorhandenen intersubjektiven Verständigung über die diagnostischen Anregungsbedingungen verfügt die Diagnostiker- und Forschergemeinschaft über keine Beobachtungssprache, die die Funktion der partiellen, empirischen Deutung von dispositionellen Eigenschaften erfüllen könnte.

Das gleiche Problem ergibt sich für die Beschreibung der Lösungsergebnisse von bearbeiteten Aufgaben und der Untersuchung der empirischen Bedeutung dieser Beschreibung. Für die diagnostische Rekonstruktion von dispositionellen Eigenschaften ist an die Klassifikation von Lösungsergebnissen die Forderung zu stellen, daß sie in einer Beobachtungssprache formuliert ist, und daß deshalb durch sie Operationen sprachlich zu repräsentieren sind, die das untersuchte Individuum in der Auseinander-

setzung mit den diagnostischen Anregungsbedingungen vollzieht. Aus der Forderung der beobachtungssprachlichen Verallgemeinerung der Operationen folgt, daß diese erlebnismäßig aufweisbar bzw. anschaulich gegeben sein müssen.

Im Beispiel des Lösungsergebnisses "Aufsatz" wurden für die Beschreibung vier Klassifikationsgesichtspunkte herangezogen. Diese Klassifikationsgesichtspunkte kennzeichnen aber Eigenschaften des fertigen Aufsatzes, die den Anforderungsvorstellungen der Beurteilergemeinschaft entsprungen sind und in keiner Weise Informationen über den Prozeß der Auseinandersetzung des Schülers mit den diagnostischen Anregungsbedingungen enthalten. Es scheint zwar plausibel zu behaupten, daß, wenn in einem Aufsatz die Erlebnisverläufe klar herausgearbeitet sind und Wiederholungen vermieden wurden, der Schüler dies angestrebt hat und wegen des gehäuftten Auftretens dieser Aufsatzmerkmale deshalb im Schüler eine entsprechende Disposition vorhanden sein muß. Das Verfahren ist aber zirkulär, solange Dispositionen auf diese Weise aus selektierten Aufsatzmerkmalen "erschlossen" werden. Es wäre dann bei einem anderen Schüler eine entsprechend gegensätzliche Disposition zu "erschließen", wenn sein Aufsatz in einer sachlichen Diktion abgefaßt ist und laufend Wiederholungen vorkommen.

Um die Aufsatzmerkmale empirisch deuten zu können und die empirische Deutung als Grundlage für die Rekonstruktion dispositioneller Eigenschaften heranzuziehen, wäre die Rekonstruktion der psychischen Prozesse erforderlich, mittels derer die diagnostische Anregungsbedingung so verarbeitet wurde, daß daraus der vorliegende, geschriebene Aufsatz resultierte. Erst bei Kenntnis dieser Prozesse - sie stellen die Transformation der diagnostischen Anregungsbedingungen durch den Schüler dar - wäre die angestrebte Aussage zu begründen, ob der Schüler unter äquivalenten Anregungsbedingungen zu äquivalenten Lösungsergebnissen gelangt. Der Begründung der Aussage entspreche dann die Einführung eines schülerspezifischen Dispositionsprädikates. Bezogen auf das Aufsatzbeispiel läßt sich behaupten, daß

die Beurteilergemeinschaft mit der praktizierten Beschreibung der Lösungsergebnisse in Prüfungssituationen über keine intersubjektiv konsistent gebrauchte Beobachtungssprache verfügt, die für eine partielle, empirische Deutung von dispositionellen Schülereigenschaften verwendet werden kann. Die Diskussion der sprachlichen Verallgemeinerung objektivierter Beurteilungsergebnisse wird zeigen, daß sich dort das gleiche Dilemma wiederholt.

Bevor ich mit der Betrachtung der Problematik der Beobachtungssprache in der diagnostischen Rekonstruktion dispositioneller Eigenschaften fortfahre, sei aus Übersichtsgründen ein zusammenfassender Kommentar eingeschoben.

2.4.3 Zusammenfassung

Die bisher geführte Diskussion über die Unterschiedlichkeit von physikalischen und diagnostischen Dispositionsprädiakten hat ergeben, daß die analogisierende Übertragung der empiristischen Einführung von Dispositionsbegriffen in die Wissenschaftssprache auf den Bereich diagnostischer Fragestellungen nicht möglich ist. Die Unterschiede äußern sich in der verschiedenen Konstitution und Struktur der jeweiligen Dispositionen. Ein darüber hinausreichender Unterschied, der die *i n t e n t i o - n a l e* Auseinandersetzung des Gegenstandes "Schüler" mit der Unterrichtssituation betrifft, wird an späterer Stelle behandelt.

Neben der Konstitutions- und Strukturdifferenz beider Dispositionsprädiakte kann man ebenfalls nach dem bisher Gesagten die unterschiedliche, beobachtungssprachliche Grundlage für die Einführung von Dispositionsbegriffen als Unterscheidungsmerkmal beider Gegenstandsbereiche ins Feld führen. Im Gegensatz zur physikalischen Beobachtungssprache, die die erlebnismäßig aufweisbar gegebene Transformation des Gegenstandes verallgemeinert, hat die Beschreibung der Verhaltensweisen von Schülern

keine entsprechende, anschauliche Voraussetzung. Es wird nicht die Transformation des Schülers als Prozeß der Lösung von Aufgaben beschrieben, sondern das Produkt bzw. Resultat des Lösungsprozesses.

Die Anwendungsbereiche der beiden Forschungsgegenstände sind also verschieden, d.h., für die formale Übertragung der empiristischen Einführung von Dispositionsbegriffen auf den Gegenstandsbereich der Diagnostik in der Schule läßt sich nach dem vorliegenden Diskussionsstand kein adäquater Anwendungsbereich festlegen und umgekehrt liegt eine Form der Einführung von Dispositionsbegriffen in die diagnostische Wissenschaftssprache nicht vor. Wenn also empiristisch orientierte Diagnostiker das empiristische Einführungsverfahren anwenden und gleichzeitig glauben, sich über ihr wissenschaftliches Vorgehen verständigen zu können, dann unterhalten sie sich über Formalisierungen und Methoden und nicht über Anwendungsbereiche. Sie unterhalten sich damit nicht über inhaltliche Interpretationen formalmethodischer Kalküle und verständigen sich lediglich über ihre gemeinsamen, instrumentellen Operationen.

Am Beispiel der idealtypischen rekonstruierten Aufsatzbeurteilung war die Verständigungsmöglichkeit vorhanden in Bezug auf die Tatsache, daß die im Lehrplan vorgesehenen Beurteilungskategorien an Erlebnisaufsätze anzulegen sind, und nicht in Bezug darauf, wie die Bedeutung der Kategorien durch die vorliegenden Aufsätze der Schüler zu konkretisieren ist. Ferner ließ sich das Verfahren intersubjektiv konsistent vermitteln, subjektiv durchgeführte Interpretationen dieser Kategorien numerisch zu symbolisieren und die Symbole rechnerisch zu syntnetisieren. Einigkeit ließ sich schließlich auch in der Berechnung der prozentualen Häufigkeit solcher Interpretationen erzielen. All dies sind formale bzw. methodische Operationen, die innerhalb der Beurteilergemeinschaft intersubjektiv konsistent eingesetzt werden können. Bei der Konstruktion und Anwendung standardisierter Testverfahren ist der entsprechende Operationenkomplex um einiges differenzierter. In dem Maße wie derar-

tige Operationen praktiziert werden, ist der gesamte Beurteilungsprozeß entsubjektiviert, d.h. von der Person des Beurteilers oder Diagnostikers unabhängig.

Das Ergebnis der Diskussion über die Frage, ob zwischen physikalischen und diagnostischen Dispositionsprädikaten hinreichende Gemeinsamkeiten existieren, um das Verfahren der empiristischen Einführung von Dispositionsbegriffen in die Wissenschaftssprache auf den diagnostischen Gegenstandsbereich übertragen zu können, bestätigt die beiden Hauptthesen dieser Arbeit:

1. Anwendungsbereich und formales Vorgehen in der praktizierten Diagnostik in der Schule sind nicht miteinander zu vereinbaren;
2. die intersubjektiv konsistente Verständigung zwischen verschiedenen Beurteilern beschränkt sich im Vollzug des Beurteilungsprozesses auf das formale Operieren mit inhaltlich unvermittelten Aussagen und Begriffen.

In der Anwendung des Formalisierungsprozesses besteht also weitgehend Einigkeit, wenn man von einzelnen syntaktischen Variationen wie zusätzliche Gewichtung von Beurteilungskategorien bei der Zensurengebung oder bestimmten Techniken der Itemselektion und der Validitätsbestimmung bei der Konstruktion von Tests absieht. Noch größere Einigkeit besteht in Bezug auf die abstrakte Forderung, den Beurteilungsprozeß zu entsubjektivieren. Die Forderung wurde in den letzten Jahren permanent als letztlisches Definitionsmerkmal einer reflektierten Forscher- und Diagnostikergemeinschaft zu proklamieren versucht, jedoch ohne entsprechende Reflexion der notwendigen Entsubjektivierungsbedingungen (INGENKAMP, 1971; INGENKAMP, 1975). In dieser vor allem für die empiristisch orientierte Forschergemeinschaft, durchgesetzten Verbindlichkeit fungiert die Entsubjektivierungsforderung als Forschungsparadigma in der von KUHN gemeinten Bedeutung, nämlich als eine scheinbar erfolgsverheißende Methode zur wissenschaftlichen Untersuchung von Gegenstandsbereichen.

Auf der anderen Seite zwingt aber die Diskrepanz zwischen der Einigkeit in der Verwendung der Formalisierungen und der divergenten Interpretation der Ergebnisse des Formalisierungsprozesses zu der Annahme, daß sich die Entsubjektivierungsforderung verselbständigt hat. Sie geht zwar als normativ wirksames Moment in die Reflexionstätigkeit von Diagnostikern bzw. Beurteilern ein und beeinflußt als internalisiertes Bewährungskriterium die Verständigungsmöglichkeit unter ihnen; sie verhindert aber zwischen ihnen den rationalen Diskurs über die den formalisierten Urteilen gegebene subjektive Bedeutung. Die Ursache dafür liegt in der nicht herstellbaren Beziehung zwischen dem Anwendungsbereich und dem formalen Vorgehen in der praktizierten Diagnostik. Bei der späteren Untersuchung der allgemeinen Funktion der Zensuren im Kontext des selektionsorientierten, organisierten Lernens wird die Ursache für das normativ wirksame Moment auf ein abstraktes Interesse der Institution "Bildung und Ausbildung" zurückgeführt.

Wenn man die bisher erfolgte Diskussion der diagnostischen Rekonstruktion dispositioneller Eigenschaften unter dem Gesichtspunkt des Entsubjektivierungsparadigmas sieht, so ist der Interpretation der KUHNschen Position durch POPPER und SCHIEFFLER zuzustimmen. In ihrer Funktion als Forscher lassen sich Diagnostiker durch eine unkritische Einstellung gegenüber und durch eine Überangepaßtheit an empiristischen Prinzipien kennzeichnen. Sie nehmen den Entstehungs- und Begründungszusammenhang solcher Prinzipien nicht zur Kenntnis und sind deshalb nicht in der Lage, deren Übertragbarkeit auf ihren Gegenstandsbereich zu überprüfen. Sie müssen deshalb auch in der Verständigung untereinander erfolglos bleiben, wenn sie die subjektive Interpretation ihrer Beurteilungsergebnisse vermitteln wollen oder eine rationale Begründung in *r e s* Beurteilungsprozesses zu geben versuchen. Gelangen sie bei der Beurteilung der objektiv gleichen Situation - z.B. der Beurteilung eines Aufsatzes - sogar zu unterschiedlichen *f o r m a l i s i e r t e n* Ergebnissen, so führen sie einen neuen Begriff zur Wiederherstellung

der verlorengegangenen Intersubjektivität ein: "fehlende Objektivität".

Der konstatierbare fehlenden Objektivität praktizierter Diagnostik in der Schule steht die komplementäre, forschungsprogrammatische Forderung nach Entsubjektivierung des Beurteilungsprozesses gegenüber. Da in der gegenwärtigen bildungspolitischen Situation die Problematisierung der praktizierten Zensurengebung ständig zunimmt und die empiristisch orientierten Wissenschaftler mit dem Versuch darauf reagieren, die Entsubjektivierungsforderung durch eine stärkere, *m e t h o d i s c h e* Objektivierung der Beurteilung einzulösen, bestätigt sich darin ein krisenhafter Zustand der diagnostischen Forschung. Vielleicht registriert man in der näheren Zukunft das Problem, daß die Entsubjektivierungsforderung in einer falschen Weise zum Forschungsparadigma geworden und daß ihre Verwirklichung zunächst noch keine *m e t h o d i s c h e* Frage ist. Folgt man der Rezeption der KUHNschen Position durch STEGMÜLLER, so ist das feststellbare Bedürfnis im Bereich der diagnostischen Forschung nach einer zunehmenden Formalisierung und Reglementierung wissenschaftlicher Tätigkeit Ausdruck des "irrationalen" Zustandes des Wissenschaftsbetriebes. Er könnte die Vorbedingung für einen zu erwartenden Paradigmenwechsel sein.

2.4.4 Die Festlegung des Beobachtbaren

Ein weiteres Argument gegen die Übertragbarkeit des Einführungsverfahrens von Dispositionsbegriffen ist eigentlich kein selbständiger, kategorialer Einwand wie die beiden vorher erörterten. Er beinhaltet die Verlängerung der Auseinandersetzung mit dem Problem der Beobachtungssprache in der Diagnostik und berührt in letzter Konsequenz den Prozeß, der im gängigen diagnostischen Sprachspiel gelegentlich als "Operationalisierung" bezeichnet wird.

Die Frage, ob den in der schulischen Diagnostik gebräuchlichen Beschreibungen von Verhaltensresultaten in Prüfungssituationen die Eigenschaft einer Beobachtungssprache zukommt, wurde bereits in spezifischer Weise beantwortet. Im Verwendungszusammenhang der diagnostischen Rekonstruktion dispositioneller Eigenschaften erfüllen derartige Beschreibungen eine bestimmte Bedingung nicht: Sie haben keine anschaulich gegebene oder erlebnismäßig aufweisbare, empirische Grundlage, d.h., sie sind keine sprachliche Verallgemeinerung der vom Schüler vollzogenen Lösungsoperationen. Mit ihnen wird lediglich eine Klassifikation der Lösung s r e s u l t a t e nach Gesichtspunkte wie "richtig-falsch", "brauchbar-unbrauchbar", "erreicht-nichterreicht", "erwünscht-nichterwünscht" usw. vorgenommen.

Vom Standpunkt der Forderung nach prognostischer Relevanz der Klassifikationen bzw. ihrer formalen Synthese für Selektionsentscheidungen ist aber zu untersuchen, inwieweit unter einem Klassifikationsgesichtspunkt zusammengefaßte Verhaltensresultate von Schülern in festgelegten diagnostischen Situationen in anderen Situationen reproduziert werden können, wo sie für die Ausführung komplexerer Lösungsoperationen erforderlich sind. Dies ist eine vor dem Hintergrund des selektionsorientierten Ausbildungswesens legitime Forschungsfrage. Mit ihrer Untersuchung soll zugleich die Antwort auf die Frage vervollständigt

werden, ob sich Aussagen über Schülerverhaltensweisen in der gleichen Art als beobachtungssprachliche Verallgemeinerungen von empirischen Operationen begreifen lassen wie die Aussagen der physikalischen Beobachtungssprache.

Eine erste, intuitive Vorüberlegung mag den Eindruck erwecken, die Fortsetzung der Problemuntersuchung sei nicht mehr nötig, da die Folge der vorgebrachten Einwände ohnehin gegen die inhaltlich-empirische Interpretierbarkeit von Eignungsbeurteilungen in der schulischen Diagnostik spricht. Demgegenüber ist aber die Tatsache zu berücksichtigen, daß das Interesse der Institution "Bildung und Ausbildung" an der prognostischen Relevanz von Schülerverhaltensklassifikationen durch die Struktur des Vorgehens in der Selektionsentscheidung erzeugt wird und deshalb als verselbständigtes Problem in Erscheinung treten kann. Wie ist die Struktur dieses Vorgehens zu kennzeichnen?

Der Unterricht in einem bestimmten Teilabschnitt des organisierten Lernens - z.B. in der Grundschule - verfolgt das Ziel, den Schülern Qualifikationen zu vermitteln, die für den Erwerb von Qualifikationen in künftigen Teilabschnitten - z.B. in den weiterführenden Schulzweigen - oder die Ausübung einer beruflichen Tätigkeit notwendige Voraussetzung sind. M.a.W.: Dem Vermittlungsvorgang innerhalb des organisierten Lernens wird eine lineare Aufeinanderfolge von Qualifikationseinheiten als seiner Struktur unterstellt. Die Sequenzierung der Qualifikationseinheiten richtet sich dabei meistens nach einer mehr oder weniger begründeten Sachstruktur des zu vermittelnden Gegenstandsbereiches. ¹²⁾.

-
- 12) Ich möchte an dieser Stelle auf die Fragwürdigkeit der in diesem Vorgehen implizierten Gleichsetzung von Aneignungsstruktur im Lerner und Sachstruktur lediglich hindeuten. BRUNERs Ansatz zur Bestimmung der Struktur der Disziplin stellt den bisher wichtigsten Versuch dar, die Struktur eines wissenschaftlichen Gegenstandsbereiches mit Aneignungsstrukturen im Lerner zum Zweck der didaktischen Rekonstruktion des Unterrichtsgeschehens zu parallelisieren (BRUNER, 1966). Die Kritik an diesem Ansatz thematisiert das Problem,

Aufgrund der unterstellten Sequenzierung wird nun behauptet, daß Schüler, die die zu vermittelnden Qualifikationen erworben haben, für eine bestimmte, weiterführende Ausbildungsalternative geeignet sind, wenn

- a) die erworbenen Qualifikationen für den Lernerfolg in der Ausbildungsalternative notwendig und
- b) in der Zeit invariant sind.

Durch die Invarianzbestimmung wird angenommen, daß jene Operationen vom Schüler in äquivalenten Lernsituationen der künftigen Ausbildungsalternative jederzeit reproduziert werden können, die für die erforderlichen Qualifikationen konstitutiv sind. Da aber - wie bereits diskutiert - unter gegebenen Prüfungsbedingungen solche Operationen diagnostisch nicht zu rekonstruieren sind und man unter den Bedingungen des selektiven Schulsystems nicht abwarten kann, ob Schüler zu einem späteren Zeit-

daß akademische Disziplinen unreflektiert zum normativen Bezugspunkt für Curricula in der Schule gemacht werden. RASCHERT (1975) weist nach, daß dieser Ansatz weitgehend ein Postulat bleiben mußte, und warum darin die Gefahr einer Reduktion des Erkenntnisprozesses von Lernern auf ausschließlich fachwissenschaftliche Strukturbestimmungen enthalten ist.

Auch die Versuche zu einer Taxonomierung von Lernzielen (BLOOM, 1956; GAGNE, 1969) gehen in letzter Konsequenz von dieser Konzeption aus, die die entwickelten Hierarchiestufen im Aneignungsprozeß des Wissens mit dem zunehmenden Komplexitätsgrad der Wissensstrukturen parallelisieren. Grundsätzlich bleibt die Frage nach der Korrespondenz der Strukturierung von wissenschaftlichen Disziplinen und Strukturierung der Lernprozesse des Lerners in der wissenschaftlichen Diskussion völlig unbeantwortet, sowohl im Hinblick auf die effektive Gestaltung von Unterrichtsprozessen als auch in Bezug auf ihre entwicklungspsychologisch begründbare Relevanz (vgl. AEBLI, 1971). Zudem ist die Aufrechterhaltung der notwendig voraussetzenden Prämisse, Wissenschaftsdisziplinen ließen sich eindeutig, intersubjektiv konsistent strukturieren, äußerst fragwürdig, wenn man von der z.T. nur partiellen Axiomatisierbarkeit formalisierter Wissenschaftsdisziplinen (Mathematik, formale Logik, u.U. Kybernetik) absieht. Und auch hier erweist sich ihre Umsetzung in didaktische Sequenzen als letztlich erfolgloses Unternehmen (FREUDENTHAL, 1974).

punkt diese Operationen in der Tat reproduzieren, muß die Stabilität bzw. Invarianz der erworbenen Qualifikationen auf andere Weise überprüft werden. Die Methode zur Feststellung der Invarianz beinhaltet zwei Bestimmungen:

1. Anstelle der Invarianz der nicht feststellbaren Operationen wird die Invarianz in der Produktion äquivalenter Verhaltensresultate oder Aufgabenlösungsprodukte überprüft. Die Invarianz ist dabei durch den Gesichtspunkt der Verhaltensklassifikation festgelegt; er wird als Qualifikation bezeichnet.
2. Der Stabilitätszeitraum wird auf die Dauer der Prüfungssituation oder auf die Dauer von mehreren Prüfungssituationen vor der Zulassung zur weiterführenden Ausbildungsalternative verkürzt.

Sofern diese Bestimmungen als erfüllt und begründet vorausgesetzt werden, kann man mit pragmatischer Plausibilität annehmen, daß alle Schüler, die in der Prüfung lernerfolgsrelevante und stabile Qualifikationen aufweisen, in der künftigen Ausbildungsalternative Erfolg haben werden.

Die pragmatische Annahme bestätigt sich in der bestehenden Schullwirklichkeit jedoch nur im statistischen Durchschnitt. Das heißt: Es gibt zwar immer wieder eine bestimmte Menge von Schülern, die sich im Sinne der Annahme verhalten, die Abweichungen davon sind aber - wiederum pragmatisch - zu stark. So erreichen z.B. im Gymnasium zu wenig zugelassene Schüler die Endqualifikation "Abitur" und bei zuvielen abgewiesenen Schülern kann man vermuten, daß sie die Endqualifikation erreicht hätten. Nicht zuletzt bilden die feststellbaren und vermuteten Abweichungen die Ursache für das Bestreben, die praktizierte Diagnostik in Richtung auf minimale Irrtumsquoten zu optimieren. Dieses Dilemma rechtfertigt die Untersuchung der Begründbarkeit der Annahme. Dabei bedürfen zwei Fragen einer Antwort, die sich im Kontext der Untersuchung der intersubjektiven Verständigung zwischen Beurteilern bzw. Diagnostikern stellen und auf die objektive

Begründbarkeit der Invarianzbehauptung abzielen:

1. Welche empirische Bedeutung können Klassifikationsgesichtspunkte zur Beurteilung von Verhaltensprodukten von Schülern haben ?

Oder unter empiristischer Perspektive: Inwieweit ist das Invariante verschiedener Verhaltensprodukte eines Schülers eine beobachtungssprachliche Verallgemeinerung der erlebnismäßig aufweisbaren Erscheinungen dieser Verhaltensprodukte ?

2. Welche Probleme ergeben sich aus der Verkürzung des vorauszusetzenden Stabilitätszeitraumes der invarianten Verhaltensaspekte auf den Zeitraum von Prüfungen ?

Oder wiederum unter empiristischer Perspektive: Läßt sich der Übergang von der situativen Invarianz auf eine langfristige als Einführung eines Dispositionsbegriffes in die Wissenschaftssprache begründen ?

Die Beantwortung beider Fragen wird im Zusammenhang der Betrachtung von Hypothesenformulierungen erfolgen, die dem implizierten Vorgehen der praktizierten Diagnostik in der Schule entsprechen und ihren naiv-empiristischen Charakter konkretisieren. Es wird sich dabei erneut zeigen, daß das diagnostische Vorgehen in der Gemeinschaft der Beurteiler die beiden Hauptthesen der Arbeit bestätigt und als Konsequenz der Orientierung am verselbständigten Entsubjektivierungsparadigma im selektiven Ausbildungssystem zu begreifen ist.

Zur Exemplifizierung der naiv-empiristischen Hypothesenformulierung gehe ich von der diagnostischen Fragestellung beim Übertritt von einer allgemeinverbindlichen Qualifizierungsphase - z.B. die Grundschule - in verschiedene weiterführende Qualifizierungsalternativen - z.B. die drei weiterführenden Schulzweige - aus. Die Fragestellung sieht in allgemeiner Form folgendermaßen aus: Läßt sich die Menge der Grundschüler auf der Basis bestimmter Vergegenständlichungsformen ihres Verhaltens in Klassen aufteilen, deren Bestimmungsmerkmale den Verbleib der

Schüler in den korrespondierenden Qualifizierungsalternativen garantieren ?

Für ihre Bearbeitung ist vorauszusetzen, das sich der Verbleib in den Alternativen durch das Erreichen angegebener Verhaltensmuster (Endqualifikationen wie etwa Abitur, Realschulabschluß oder qualifizierter Hauptschulabschluß) kennzeichnen läßt. Ferner ist vorauszusetzen, daß Diagnostiker aufgrund von Erfahrungen und von in der Wissenschaft gesammelten Befunden über Vermutungen verfügen, welche Eingangsqualifikationen für die verschiedenen Qualifizierungsalternativen notwendig vorausgesetzt werden müssen, damit die Endqualifikationen durch die Schüler zu erreichen sind. Auf der Grundlage solcher Vermutungen wäre es scheinbar möglich, empiristische Hypothesen über eine pädagogisch sinnvolle Klassifikation der Grundschüler zu bilden, und zwar von folgender Art:

- Wenn jemand die Differentialrechnung beherrscht, dann beherrscht er auch die vier Grundrechnungsarten.
Oder negativ formuliert:
Wer mit den Grundrechnungsarten nicht umgehen kann, wird mit der Differentialrechnung nicht umgehen können.
- Wenn jemand einen Besinnungsaufsatz schreiben kann, dann verfügt er auch über den richtigen Gebrauch der syntaktischen Regeln.
Negativ formuliert:
Wer die syntaktischen Regeln nicht richtig gebraucht, ist nicht in der Lage, einen Besinnungsaufsatz zu schreiben.
- Wenn jemand einen fremdsprachlichen Text richtig übersetzt, dann verfügt er auch über einen umfangreichen Wortschatz.
Negativ formuliert:
Wer keine umfangreichen Wortschatzkenntnisse hat, scheitert an der Übersetzung eines fremdsprachlichen Textes.

Den genannten Beziehungen ist gemeinsam, daß in der positiven Formulierung im Dann-Satz notwendige Voraussetzungen in Form von Qualifikationen angegeben werden (aus Demonstrationsgründen habe ich nur jeweils eine solche notwendige Voraussetzung unter realiter vielen erwähnt). Unabhängig von den Schwierig-

keiten, im konkreten Fall die Gesamtheit der notwendigen Voraussetzungen zu bestimmen, verbirgt sich das zentrale Problem hinter dem Charakter der Beziehungen, wenn man dazu übergeht, das Vorhandensein der notwendigen Qualifikationen zu erfassen und das Ergebnis der Diagnose zur Klassifikationsgrundlage der Grundschüler macht.

Denn, in den drei genannten Beziehungen wird zunächst nur ausgesagt, daß zum Vollzug komplexer Operationen wie "Differenzieren", "Schreiben eines Besinnungsaufsatzes" und "Übersetzen eines Textes" die Ausführung von Teiloperationen notwendig ist, ohne die die Durchführung der komplexen Operationen nicht gelingt. Aus diesem Grund habe ich vorher von der scheinbaren Möglichkeit der Formulierung von empiristischen Hypothesen gesprochen. Die genannten Beziehungen haben zwar die Form von Hypothesen, da in ihnen ein Zusammenhang von Phänomenen behauptet wird. Der behauptete Zusammenhang in der gegebenen Formulierung ist allerdings kein empirischer: So ist z.B. die Behauptung, daß im Vollzug der Differentialrechnung die vier Grundrechnungsarten zur Anwendung kommen immer und unabhängig von der ausführenden Person richtig.

Sie durch die Feststellung der Tatsache "empirisch" überprüfen zu wollen, daß alle Personen, die die vier Grundrechnungsarten nicht beherrschen, auch die Differentialrechnung nicht ausführen können, ist daher unsinnig. Das gleiche gilt auch für die beiden anderen Beziehungen, sofern die darin vorkommenden Ausdrücke hinreichend präzisiert sind. Daß die drei formulierten Beziehungen als empirische Hypothesen erscheinen können liegt daran, daß, obwohl sie analytische Beziehungen zwischen abstrakten Operationen benennen, Zusammenhänge zwischen Eigenschaften von konkreten Objekten - hier die Schüler - formulieren sollen. Da die abstrakten Operationen als allgemeine Fähigkeiten von Individuen erscheinen, wird die Analytizität der Beziehungen verschleiert.

Die "operationalisierende" Veränderung des Charakters der Beziehungen bildet für sich genommen kein Problem. An der Gültigkeit

der analytischen Beziehungen ändert sich nichts, wenn man Personen die Operationen ausführen läßt. Die scheinbar empiristische Überprüfung der Aussagen erbringt keinerlei Bedeutungszuwachs. M.a.W.: Die drei Aussagen sind nicht auf Aussagen zurückzuführen, die die ersteren im empiristischen Sinne zumindest partiell empirisch interpretieren. Solche partiell interpretierenden Aussagen sind aber bei der Einführung von Dispositionsbegriffen in die Wissenschaftssprache Aussagen der Beobachtungssprache.

Nehmen wir an, man wollte durch die drei Aussagen diagnostische Dispositionsbegriffe in die Wissenschaftssprache einführen - durch die erste Aussage eine Disposition für das Fach Mathematik, durch die zweite für das Fach Deutsch und durch die dritte für ein fremdsprachliches Fach; nehmen wir weiterhin an, konkrete Schülerverhaltensweisen in Prüfungssituationen würden z.B. unter dem Gesichtspunkt "Beherrschen der vier Grundrechnungsarten" und "Beherrschen der Differentialrechnung" klassifiziert. Unter dieser Voraussetzung wäre die Feststellung, daß die Verhaltensprodukte einer bestimmten Menge von Schülern zum einen gelöste Aufgaben aus dem Bereich der vier Grundrechnungsarten und zum anderen aus dem Bereich der Differentialrechnung darstellen (die Feststellung erfolgt natürlich zu verschiedenen Zeitpunkten), keine empiristische Bestätigung der Aussage "Wenn jemand die Differentialrechnung beherrscht, dann beherrscht er auch die vier Grundrechnungsarten".

Die Aussage beschreibt nur ein Strukturmerkmal der Differentialrechnung. Daß sie in der oben beschriebenen Weise *s c h e i n - b a r* zu überprüfen ist, d.h., daß zu unterschiedlichen Zeitpunkten die Beherrschung der vier Grundrechnungsarten für sich und die Beherrschung der Differentialrechnung inklusive der vier Grundrechnungsarten festzustellen ist, liegt am Aufbau des organisierten Lernens und dem darin praktizierten Selektionsmodus. Es gelangen nämlich nur jene Schüler in eine Selektionsalternative mit Ausbildung in Differentialrechnung, die die vier Grundrechnungsarten zum früheren Zeitpunkt beherrschen.

Die empirische Bedeutungslosigkeit der Aussagen äußert sich somit in der tautologischen Bestätigungsprozedur. Aus diesem Grund kann von einer Einführung eines diagnostischen Dispositionsbegriffes auf der Basis solcher Bestätigungsverfahren keine Rede sein. Daraus sind zwei alternative Konsequenzen zu ziehen:

1. Man begreift die klassifizierende Beschreibung von Schülerverhaltensweisen als eine Art beobachtungssprachlicher Verallgemeinerung. Dann muß man den Anspruch auf Rückführbarkeit von Dispositionsbegriffen auf diese Beobachtungssprache preisgeben. Mit Aussagen der eingangs zitierten Art wird dann nur die Folge der Vermittlung von Lehrinhalten beschrieben, wie sie einer angenommenen Disziplinstruktur entspricht.
2. Man stellt fest, daß es sich bei der klassifizierenden Beschreibung von Schülerverhaltensweisen um keine Beobachtungssprache handelt, die für die Einführung von Dispositionsbegriffen notwendig ist. Dann muß untersucht werden, welche Bedeutungen solche Beschreibungen haben und wie sie in die praktizierte diagnostische Rekonstruktion von dispositionellen Eigenschaften Eingang finden.

Die erste Konsequenz bedarf keiner Diskussion, weil sie in Bezug auf die Festlegung dessen, was unter einer Beobachtungssprache in der diagnostischen Begriffsbildung zu verstehen ist, Willkürlichkeit unterstellt. Die zweite Konsequenz muß diskutiert werden, wenn man sie im Kontext der selektionsrelevanten Hypothesenbildung für die diagnostische Rekonstruktion dispositioneller Eigenschaften betrachtet. Dazu ist es allerdings notwendig, die Funktion der drei scheinhypothetischen Aussagen für die diagnostische Begriffsbildung aufzuzeigen.

Bei der konkreten diagnostischen Forschungsarbeit ist die empiristische Bestätigung von Scheinhypothesen in der angegebenen Weise erst gar nicht vorgesehen. Zum Zweck der Gewinnung prognostisch relevanter Beurteilungsinformationen geht man von vornherein von der umgekehrten Fragestellung aus. Man benutzt da-

bei die negative Formulierung der Scheinhypothesen, die ich unter der positiven Formulierung jeweils aufgeführt habe und die in allgemeiner Form folgendermaßen aussieht:

- Wer die notwendigen Operationen nicht beherrscht, der kann die komplexen Operationen nicht beherrschen.

Obwohl diese Form logisch äquivalent ist mit der positiven Formulierung und deshalb die gleichen Bestätigungsprobleme aufwirft, suggeriert sie eine bessere Brauchbarkeit für die Lösung des Dispositionsproblems. Ihre Selektionsfunktionalität tritt fast anschaulich hervor. Denn unter der Voraussetzung selektionsstrukturierter Ausbildung, nach der in eine Selektionsalternative nur zugelassen werden sollte, wer zum Übertrittszeitpunkt über die notwendigen Qualifikationen verfügt, läßt sich die Gesamtmenge der Bewerber nach scheinbar prognostisch relevanten Kriterien in zwei Klassen aufteilen: in eine Klasse mit Schülern, die für die Selektionsalternative in Frage kommen, weil sie über die notwendigen Voraussetzungen verfügen, und in eine Klasse mit aussichtslosen Kandidaten.

Auf der anderen Seite tritt die darin enthaltene Bestätigungsproblematik ebenso deutlich hervor. Die Nichtzulassung von unzureichend qualifizierten Schülern mag in anbetracht der Anforderungsbedingungen in der aufnehmenden Qualifizierungsalternative noch einigermaßen pragmatisch legitimierbar sein. Man kann sich auf den Standpunkt stellen, daß die Defizite der unzureichend qualifizierten Schüler zu groß sind, um sie durch den Unterricht in der aufnehmenden Qualifizierungsalternative und in der zur Verfügung stehenden Zeit noch kompensieren zu können. Dieses pragmatische Argument ist sogar konstitutiv für die Interpretation von Zensuren. Nach dem Beschluß der ständigen Konferenz der Kultusminister der Länder und in der Bundesrepublik Deutschland vom 3.10.68 wird die Note "mangelhaft" erteilt, wenn neben anderen Bedingungen die Mängel in absehbarer Zeit behoben werden können, und die Note "ungenügend", wenn die Mängel in absehbarer Zeit nicht behoben werden können (KMK, 1969). Unüberwindbare Schwierigkeiten dagegen bereitet die Begründung

für die Zulassung von Schülern, die über die vorausgesetzten Qualifikationen verfügen. Da die Qualifikationen einerseits nur notwendige Voraussetzungen sind und andererseits ihre prognostische Relevanz erst nachzuweisen ist, reicht ihre Bedeutung als Informationsgrundlage für zuverlässige Selektionsentscheidungen nicht über eine erwartungstreue Hoffnung hinaus.

Für den Nachweis der prognostischen Relevanz von Qualifikationen ist zu überprüfen, ob die unter dem Qualifikationsgesichtspunkt klassifizierten Verhaltensprodukte von Schülern in Prüfungssituationen langfristig stabil bleiben, d.h. unter strukturell gleichen Bedingungen reproduziert werden können. Auch wenn die Überprüfung der Stabilität über den Zeitraum der potentiellen Qualifizierungsalternative nicht möglich ist und zusätzliche, theoretische und methodische Probleme aufwirft, so muß doch begründet werden, wie die Annahme der Stabilität auf der Basis von Urteilen über Verhaltensprodukte aufrechtzuerhalten ist.

Der erste Schritt auf dem Weg zur Rechtfertigung der Annahme besteht darin, daß aus der wiederholten Herstellung gleicher Lösungsprodukte durch den Schüler auf stabile bzw. invariante Merkmale seiner Persönlichkeit geschlossen wird. Die dabei implizierte Identifizierung der Gleichheit von Lösungsprodukten mit der unterstellten Gleichheit von Lösungsprozessen wurde bereits problematisiert. Der zweite Schritt bezieht sich auf die Festlegung der Gleichheit von Lösungsprodukten. Es stellt sich hier die Frage, wie ein Beurteiler dazu kommt, einzelne Lösungsprodukte von Schülern als gleich oder ungleich anzusehen. Sie tritt im Bereich der objektivierten Diagnostik in der Schule als Forderung nach einer Meßtheorie in Erscheinung. Ich möchte aber die Untersuchung der Frage auf klassifikatorische Begriffe beschränken, weil die Einführung von höheren Begriffsformen wie z.B. die metrischen, die Verwendung eines Meßinstruments voraussetzen. Dies gilt auch für komparative Begriffe, wenn sie intersubjektiv konsistent benutzt werden

sollen. Für die Problematisierung der Beurteilung gleicher Lösungsprodukte ist die Berücksichtigung der Beziehung zwischen der Einführung eines metrischen Begriffes, der Verfügung über ein Meßinstrument und der Messung nicht erforderlich.

Allgemein verfolgt man mit klassifikatorischen Begriffen den Zweck, Gegenstände eines Bereiches - hier des Bereiches der Lösungsprodukte eines Schülers in Prüfungssituationen - in verschiedene Klassen zu zerlegen. Die Klassifikation muß nach STEGMÜLLER (1970, Band II, Studienausgabe Teil A) zwei Adäquatheitsbedingungen erfüllen:

"(I) Die durch die einzelnen Begriffe festgelegten Klassen müssen gegeneinander scharf abgegrenzt sein ... Die einzelnen Klassen der Einteilung müssen sich wechselseitig ausschließen.

(II) Auf der anderen Seite muß auch jeder Gegenstand des Bereiches in eine der Klassen fallen, die durch die Begriffe festgelegt sind, genauer ausgedrückt: die Klasseneinteilung muß erschöpfend sein, so daß jedes Objekt des Bereiches unter eine der begrifflich festgelegten Klassen fällt" (ebd., S. 19).

Es gibt nun zwei grundsätzlich verschiedene Fälle, in denen die Klassifikation eines Gegenstandsbereiches die beiden Adäquatheitsbedingungen erfüllt: Im ersten Fall ist die Erfüllung der Bedingungen eine logische Folgerung der Definitionen der Klassifikationsbegriffe. Beispielsweise liegt eine ausschließlich definitorische Klassifikation vor, wenn man für den Umgang mit den vier Grundrechnungsarten festlegt, was eine Richtiglösung und was eine Falschlösung ist. Dies läßt sich prinzipiell auch für den Gebrauch syntaktischer Regeln oder für die Umfangsbestimmung des verfügbaren Wortschatzes festsetzen, wenn auch nicht in der gleichen Eindeutigkeit. Bei hinreichender Reglementierung der Festsetzung könnte man jedenfalls erreichen, daß die Verhaltensresultate von Schülern in Prüfungssituationen in zwei sich jeweils wechselseitig ausschließende und erschöpfende Klassen aufteilbar sind.

Der zweite Fall geht über die Festsetzungsebene hinaus. Hier

wird die Forderung erhoben, daß die Klassifikationsmerkmale empirische Bedeutung haben. Unter dieser Voraussetzung "drückt entweder die Aussage, daß die Begriffe einander ausschließen, oder die Aussage, daß ihre Umfänge den gesamten Bereich erschöpfen, oder daß beides gilt, keine logische Wahrheit, sondern **e i n N a t u r g e s e t z a u s**" (ebd., S. 22). STEGMÜLLER exemplifiziert die Erfüllung beider Bedingungen an der Klassifikation "männlich-weiblich" und "bunt-unbunt". Er zeigt auf, daß die Bedingungen in diesem Fall nur dann zu erfüllen sind, wenn die einzelnen Klassenbegriffe **u n a b h ä n g i g v o n e i n a n d e r** durch empirische Befunde festgelegt werden, die sich in Form einer empirischen Gesetzesaussage generalisieren lassen. Am Beispiel der Unterscheidung "männlich-weiblich" sind das einige biologische Merkmale, die unabhängig voneinander an den konkreten Objekten "Mann" bzw. "Frau" als vorfindbare ausgewählt wurden. Zwar gehen auch in den empirisch begründeten Klassifikationsvorgang durch die Merkmalsauswahl Festsetzungen ein, die Klassifikation der Gegenstände ist aber nicht analytisch gültig, d.h., es kann nicht gesagt werden, "was in die Klasse A fällt, gehört nicht in die Klasse B und umgekehrt".

Die Klassifikation ist damit grundsätzlich empirisch widerlegbar, z.B. durch das Finden von Zwittern. Die Widerlegbarkeit ist eine Konsequenz der Konstitution des Klassifikationsprozesses. Die Klassifikationsmerkmale wurden in logischer Unabhängigkeit voneinander **u n e** an den betrachteten Objekten gewonnen, sei es durch unmittelbare Beobachtung oder durch experimentell-instrumentellen Zugang. Infolgedessen ist das Ergebnis der Klassifikation von den betrachteten Objekten mitabhängig und kann durch andere Objekte falsifiziert werden.

Letzteres ist in der Klassifikation im Bereich der praktizierten Diagnostik in der Schule nicht gegeben. Unter der Voraussetzung, hinreichender Standardisierung des Beurteilungsprozesses werden die Verhaltensresultate von Schülern in Prüfungssituationen in dichotome, logisch voneinander abhängige Klassen eingeteilt. Doren Festsetzung wird zwar zum Gegenstand der

Diskussion über intersubjektive Konsistenz, sie hat aber immer die Endform einer mehr oder weniger vereinheitlichten und verbindlich interpretierten Polarisierung von kontradiktorischen Begriffen wie "richtig - nicht richtig", "erwünscht - nicht erwünscht", "gelöst - nicht gelöst", "brauchbar - nicht brauchbar" usw. M.a.W.: Da die Klassifikationsmerkmale zueinander in einem kontradiktorischen Gegensatz stehen und eines davon in jedem Fall einem gegebenen Lösungsprodukt zugeordnet wird, muß die Aussage "eine Aufgabenlösung ist richtig o d e r nicht richtig, erwünscht o d e r nicht erwünscht, usw." immer richtig sein und schafft die Voraussetzung dafür, ausschließlich durch Festlegung die positive Alternative zu bestimmen.

Ob die positive Bestimmung empirisch bedeutungsvoll ist oder nicht, läßt sich nicht entscheiden, da es immer Lösungsprodukte gibt, auf die eine der beiden Beurteilungspole zutrifft.

Die "negativen" Problemlösungen bestimmen sich daher immer nur durch ihr Verhältnis zu den positivierten und sind nur als Abweichungen von den Sollvorstellungen zu interpretieren, die durch die Klassifikationsmerkmale festgesetzt sind.

Trotz allen fehlenden, empirischen Bedeutungsgehalts wird aus diesem Klassifikationsvorgehen für die Annahme der Stabilität bzw. Invarianz von Schülerverhaltensweisen auf der Grundlage von Urteilen über Verhaltensprodukte eine fatale Konsequenz gezogen: Je mehr Verhaltensprodukte eines Schülers in Prüfungssituationen nach Gesichtspunkten klassifiziert werden können, unter denen sie als gleiche erscheinen, desto mehr kann man scheinbar annehmen, daß im Schüler ein dispositioneller Komplex von Lösungsoperationen wirksam ist. Ihre Gleichheit kann zwei Formen haben: sie können mit dem gewünschten Lösungsprodukt übereinstimmen oder nicht übereinstimmen.

Es sieht im ersten Moment so aus, als wäre man jetzt genötigt, zwei Arten von Dispositionen anzunehmen, eine Disposition zur Problemlösung und eine andere, die man als Unfähigkeit zur Problemlösung bezeichnen müßte. Da aber beide Dispositionen in einem kontradiktorischen Verhältnis zueinander stehen, läßt sich

die "negative" Disposition nicht unabhängig von der positiven interpretieren. Aus diesem Dilemma wird ein pragmatischer Ausweg gewählt: Man faßt die Prüfungssituation als diagnostische Anregung einer bestimmten Disposition auf und interpretiert die Menge der Abweichungen der erzielten Lösungsprodukte von den Sollvorstellungen als reduzierten Ausprägungsgrad dieser Disposition. Auf diese Weise wird die Menge der erwünschten Problemlösungen zum Indikator für den Ausprägungsgrad einer Disposition, deren Inhaltlichkeit durch die ausschließliche analytischen Bestimmungen festgelegt ist.

Setzt man voraus, daß die Bedeutung einer Beobachtungssprache in der Verallgemeinerung empirischer Phänomene besteht - seien diese in der anschaulichen Erfahrung unmittelbar gegeben oder durch Beobachtungsinstrumente vermittelt -, dann muß der Beschreibung der Verhaltensresultate von Schülern in Prüfungssituationen eine beobachtungssprachliche Eigenschaft abgesprochen werden. Darüber sollte die Tatsache nicht hinwegtäuschen, daß die Klassifikationsmerkmale von Lösungsprodukten für den Beurteiler unmittelbar feststellbar sind. Sie wurden nicht nachträglich aus dem Vergleich vorhandener Lösungsprodukte gewonnen, sondern lagen bereits vor dem diagnostischen Prozeß als Beurteilungsraster vor; und zwar als Sollvorstellungen über die geforderte Qualität des Lösungsproduktes. Nur dadurch ist es ja erst möglich, die Frage zu stellen, ob die produzierten Lösungen eines Schülers den Sollvorstellungen entsprechen oder nicht, d.h., ob der Schüler dazu tendiert, Lösungen zu produzieren, die unter einem bestimmten Gesichtspunkt als Reproduktionen zu klassifizieren sind oder nicht, bzw. ob der Schüler über eine dispositionelle Eigenschaft verfügt oder nicht.

Hier zeigt sich die Umkehrung der Methode bei der Einführung von Dispositionsbegriffen in die Wissenschaftssprache in der Funktion der begrifflichen Verallgemeinerung von Phänomenen. Aus dieser Perspektive soll der dritte Einwand gegen die Übertragbarkeit des Einführungsverfahrens von Dispositionsbegriffen zusammengefaßt werden.

Zusammenfassung

Bei der empiristischen Einführung von Dispositionsbegriffen ging es um das Problem, inwieweit die begriffliche Verallgemeinerung der selektiven Betrachtungsweise gegenüber dem Gegenstand und seiner Veränderungen entspricht. Die selektive Betrachtung des Gegenstandes wird deshalb mehrmals vorgenommen um zu überprüfen, ob die begriffliche Verallgemeinerung haltbar ist. Wenn die Beschreibung das Invariante des empirischen Geschehens bei allen betrachteten Objekten des Gegenstandsreiches verallgemeinert, dann hat sie die Eigenschaft einer Beobachtungssprache.

Bei der diagnostischen Rekonstruktion dispositioneller Eigenschaften geht es darum, ob für eine vorgegebene, begriffliche Verallgemeinerung einer selektiven Betrachtungsweise Objekte gefunden werden können, die die Betrachtungsweise dann partiell bestätigen sollen. Man interessiert sich also nicht für ein gemeinsames Merkmal aller betrachteten Objekte - hier der Verhaltensprodukte -, sondern für die Möglichkeit, Objekte zu finden, die unter verbindlich durchgesetzte Sollvorstellungen zu subsumieren sind. Dieser Prozeß wird in der naiv-empiristischen, pädagogischen Literatur fälscherlicherweise als "Operationalisierung" bezeichnet (MÖLLER, 1969; MAGER, 1971). Das Problem dabei ist, daß hier nicht ein bereits empirisch bedeutungshaltiger und theoretisch funktionaler Begriff durch Einführungsregeln in seiner wissenschaftssprachlichen Funktion festgelegt werden soll, sondern ein in seiner empirischen Bedeutung unklarer Begriff einer empirischen Interpretation bedarf. Da aber der Interpretationsrahmen ein analytischer ist, läßt sich die Menge und der Umfang der darunter subsumierbaren, empirischen Objekte nicht eingrenzen. So wird es immer eine beliebige Anzahl von scheinbar beobachtungssprachlich formulierten Hypothesen geben, die als "Operationalisierungen von Lernzielen" aufgefaßt werden und formale Zusammenhänge zwischen diagnostischen Anregungsbedingungen und Verhaltensprodukten formulieren; es wird auch immer Schüler geben, für die der formulierte Zu-

sammenhang zutrifft, und andere, die die Hypothese "widerlegen". Denn das Invariante an verschiedenen Verhaltensprodukten bestimmt sich nicht nach ihren gemeinsamen, empirischen Eigenschaften, sondern nach der davon unabhängigen Betrachtungsweise des Beurteilers. Und nach dieser Betrachtungsweise sollen solche Hypothesen zur Diskriminierung von Schülern verhelfen, d.h., es soll Schüler geben, die die Hypothese "bestätigen", und solche, die es nicht tun.

Die Frage, inwieweit das Invariante verschiedener Verhaltensprodukte eines Schülers eine beobachtungssprachliche Verallgemeinerung der erlebnismäßig aufweisbaren Erscheinungen dieser Verhaltensprodukte ist, muß negativ beantwortet werden. Zugleich hat sich bei der Diskussion der Frage gezeigt, daß die Aufrechterhaltung der Stabilitäts- bzw. Invarianzannahme nur durch die Scheinquantifizierung der Verhaltensprodukte zu retten ist. Da man die mit den Sollvorstellungen übereinstimmenden Verhaltensprodukte in ihrer Gleichheit nicht als empirisch bedeutungshaltig interpretieren und sich um eine davon unabhängige Interpretation der nicht übereinstimmenden Verhaltensprodukte nicht bemühen kann, muß man sich auf ihre Zählbarkeit als einzige, verbleibende Invariante zurückziehen. Was im Rekonstruktionsmodell für die Aufsatzbeurteilung als Folge der Selektionsbezogenheit der Schülerbeurteilung veranschaulicht wurde, tritt hier als naiv-empiristische Rekonstruktion dispositioneller Eigenschaften in Erscheinung: es vollzieht sich ein nichtbewußter Übergang von klassifikatorischer Beurteilung zu scheinquantifizierender Messung.

Die Zählung von als gleichartig klassifizierten Verhaltensprodukten wird somit zur Legitimation für die Bemessung des Ausprägungsgrades dispositioneller Eigenschaften. Zu den ungelösten Problemen, die mit der Einführung von diagnostischen Dispositionsbegriffen in die Wissenschaftssprache verbunden sind, treten die Bedingungen für die Einführung von quantitativen Begriffen hinzu. Auf sie kann in extenso nicht eingegangen werden. Da aber die Beurteilung in der praktizierten Diagnostik in der

Schule so gehandhabt wird, als lägen quantitative Begriffe vor, und da den Ausprägungsgraden der erfaßten Schülermerkmale prognostische Relevanz zukommen soll, ist das wesentliche in der Einführung von quantitativen Begriffen im Hinblick auf ihre empirische Interpretierbarkeit zu skizzieren. Durch den Vergleich der Anforderungen an die Einführung von quantitativen Begriffen in die Wissenschaftssprache mit dem naiven Rechtfertigungsvorgehen in der praktizierten Diagnostik in der Schule wird die Unbegründbarkeit der zeitlichen Verkürzung aufzuweisen sein, die im Beurteilungsvorgehen in Form von Prüfungen durchgesetzt ist. In diesem Kontext muß dann auch ein letztes Dilemma diagnostischer Begriffsbildung besprochen werden: die stillschweigende Ersetzung der intraindividuellen Invarianz im Aufgabenlösungsverhalten durch die interindividuelle Invarianz in der beurteilten Schülerpopulation.

Für die Weiterführung der Betrachtung sollen die bisherigen Einwände gegen die Übertragbarkeit des empiristischen Einführungsverfahrens in den Bereich der diagnostischen Rekonstruktion dispositioneller Eigenschaften und damit gegen die empirische Interpretierbarkeit praktizierter Diagnostik in der Schule im Überblick zusammengefaßt werden:

1. Im Gegensatz zum zweistelligen Dispositionsbegriff, dessen Einführung in der empiristischen Wissenschaftstheorie diskutiert wird, handelt es sich beim diagnostischen um einen dreistelligen. In der Diagnostik versucht man Aussagen über Subjekte zu machen, die sich mit einer vorgegebenen Anregungssituation auseinandersetzen und das Ergebnis der Auseinandersetzung in Form von vergegenständlichten Verhaltensprodukten äußern. Die dabei im Subjekt stattfindenden Transformationsprozesse zu erfassen, ist das eigentliche Untersuchungsziel.
2. Die Transformationsprozesse sind dem Diagnostiker prinzipiell nicht zugänglich. Sie sind weder anschaulich gegeben noch über Untersuchungsinstrumente zu vermitteln - eine Vermittlung könnte hier lediglich durch die Untersuchung organischer Prozesse, insbesondere der zentralnervösen, geschaffen werden (vgl. SCHMIDT, 1974). Alle Bemühungen, die Beurteilung durch Instrumente zu objektivieren, müssen von der Beziehung Anregungsbedingung-Subjekt-Verhaltensprodukt ausgehen und damit von der Dreistelligkeit des einzuführenden Dispositionsbegriffes.
3. Ziel der Beurteilung ist die Ermittlung von invarianten Ver-

haltensmustern bzw. invarianten Transformationsprozesse im untersuchten Individuum. Aufgrund von 2. muß deshalb versucht werden, solche Verhaltensmuster aus dem wiederholten Auftreten von gleichen Verhaltensprodukten unter strukturell gleichen Anregungsbedingungen zu erschließen. Dazu wäre ein empirisch bedeutungshaltiger Vergleichsmaßstab für verschiedene Anregungsbedingungen und Verhaltensprodukte erforderlich. Die verwendeten Vergleichsmaßstäbe in der praktizierten Diagnostik in der Schule sind aber ausschließlich definitorisch festgelegt und lassen keine empirische Deutung zu.

2.4.5 Die Standardisierung der Schülerbeurteilung als Ersatz für ihre empirische Begründung

Nach dem vorliegenden Diskussionsergebnis erweist sich das Vorgehen in der praktizierten Diagnostik als unvereinbar mit den Ansprüchen einer empiristischen Methodologie zur Gewinnung von interpretierbaren Dispositionsbegriffen. Die Hauptursache für die Unvereinbarkeit liegt in der Verdoppelung des Anwendungsbereiches begründet, über den man in der Diagnostik in der Schule Aussagen treffen will.

Einerseits ist für die diagnostische Rekonstruktion dispositioneller Schülereigenschaften der Bereich vergegenständlichten Verhaltens die Grundlage. Das Ziel der Rekonstruktion besteht in der Gewinnung von prognostisch relevanten Invarianzaspekten von Verhaltensprodukten in Prüfungssituationen. Der Zusammenhang von Prüfungssituation als diagnostische Anregungsbedingung und Verhaltensprodukten konstituiert die erste Art eines verselbstständigten Anwendungsbereiches. Diagnostische Anregungsbedingungen und Verhaltensprodukte sind die objektive Voraussetzung für die Rekonstruktbildung.

Auf der anderen Seite sollen aber die als invariant festgestellten Zusammenhänge zwischen den Voraussetzungen als Ausdruck invarianter Verarbeitungsprozesse im Schüler interpretiert werden, um die Vorhersage der Wiederholbarkeit solcher Verarbeitungsprozesse zu gestatten. Die Rekonstruktion invarianter Verarbeitungsprozesse im Schüler ist der zweite verselbständigte Anwendungs-

bereich diagnostischer Untersuchungen. Objektive Voraussetzung für die Rekonstruktbildung ist hier der Schüler, d.h. die Tatsache, daß die festgestellten Zusammenhänge zwischen diagnostischen Anregungsbedingungen und Verhaltensprodukten nicht ohne Rekurs auf den Schüler erklärt werden können.

Beide Anwendungsbereiche erscheinen im Forschungsprozeß als verselbständigte, weil ihre objektiven Voraussetzungen innerhalb der gleichen Untersuchungssituation unterschiedlich rekonstruiert werden: Für den Vergleich der Verhaltensprodukte verwendet man festgelegte, klassifikatorische Begriffe, für den Vergleich der Schüler das Ergebnis der rechnerischen Synthese von Ziffern, die die Klassifikationsmerkmale der Verhaltensprodukte symbolisieren. Letztere haben deshalb die Form von quantitativen Begriffen.

Abgesehen von der Tatsache, daß beide Begriffsformen die empiristisch-wissenschaftstheoretische Forderung der empirischen Deutbarkeit nicht erfüllen, führt die Verdoppelung des Anwendungsbereiches zu einem tautologischen Begründungsversuch für die Annahme von dispositionellen Eigenschaften. Mit diagnostischen Hypothesen über den Zusammenhang zwischen Anregungsbedingungen und Verhaltensprodukten wird sowohl der Vergleich zwischen den Verhaltensprodukten als auch der Vergleich zwischen den Schülern zugleich definitorisch eingeführt und zu bestätigen versucht. Werden beide Vergleiche mittels einer derartigen diagnostischen Hypothese in einem Durchgang vollzogen, so hat man sich daran gewöhnt, von dem Vorhandensein einer dispositionellen Eigenschaft zu sprechen. Die Verwendung des Dispositionsbegriffes scheint dann legitimiert zu sein, wenn Schüler dargebotene, äquivalente Prüfungsaufgaben wiederholt lösen und wenn die Häufigkeit der Wiederholungen zwischen den Schülern unterschiedlich ist.

Auf diese Weise werden für die Beurteilung von Schülern von vornherein nur diskriminierende Kategorien zugelassen, die aber, weil sie die diskriminierende Funktion erfüllen, als Ausdruck objektiver Dimensionen in der Schülerpersönlichkeit aufgefaßt

werden. Das, was also für die empiristische Rekonstruktion dispositioneller Eigenschaften von konkreten, einzelnen Schülern vorauszusetzen wäre, nämlich ein empirisch bedeutungsalter Maßstab zur Unterscheidung von Schülern, wird durch die Realisierung des Verfahrens festgelegt und bestätigt sich deshalb erwartungsgemäß.

Im Gegensatz dazu unterscheidet man in der empiristischen Wissenschaftstheorie eindeutig zwischen der Einführung von Dispositionsbegriffen und der Einführung von quantitativen Begriffen. Außerdem ist die Form der Einführung von Dispositionsbegriffen dort für einen Anwendungsbereich vorgesehen, innerhalb dessen nur ein Untersuchungsgegenstand betrachtet wird und zwar derjenige, an dem sich die Transformationsprozesse vollziehen und in naturgesetzlicher Form ablaufen. Eine Unterscheidung zwischen einem sich transformierenden Gegenstand und einem davon getrennt zu untersuchenden und vergegenständlichtem Produkt findet nicht statt.

Gleichwohl ist die Formalisierung der Schülerbeurteilung nach diesem Verfahren ihr allgemeines Strukturmerkmal. Im Rekonstruktionsmodell für die Aufsatzbeurteilung wurde dies idealtypisch nachvollzogen. Bei der Konstruktion und Anwendung von standardisierten Beurteilungsverfahren liegt das gleiche Verfahren vor. Dort werden allerdings die zu verrechnenden Klassifikationskategorien nicht erst nachträglich aus dem vorliegenden Prüfungsergebnis rekonstruiert, sondern vorher in einzelnen, nach dem Prinzip maximaler Diskrimination entwickelten Testaufgaben konkretisiert. Darin ist der wesentliche Unterschied zwischen objektivierter Schülerbeurteilung und konventioneller Zensurengebung zu sehen. Dies dürfte auch der entscheidende Grund aus naiv-empiristischer Sicht sein, standardisierte Beurteilungsverfahren als "objektiv" zu bezeichnen.

2.4.5.1 Die "Vorzüge" der objektivierten Schülerbeurteilung gegenüber der Zensurengebung

Obwohl auch bei der konventionellen Zensurengebung so vorgegangen wird, daß man Verhaltensprodukte von Schülern in Prüfungssituationen klassifiziert und die numerische Symbolisierung der Klassifikationsmerkmale rechnerisch synthetisiert, ist ihre größere intersubjektive Inkonsistenz zunächst auf die Festlegung und den Deutungspielraum der Klassifikationsgesichtspunkte durch und für den individuellen Lehrer zurückzuführen. In diesem Punkt unterscheiden sich die Bewertungen von Aufsätzen und die Bewertungen von Mathematikarbeiten oder anderen Arbeiten prinzipiell nicht.

Bei der Beurteilung von Aufsätzen bzw. von Übersetzungen, Geschichtsexamensergebnissen u.ä. ermöglicht zwar der semantische Informationsgehalt der Verhaltensprodukte spezielle, subjektivistische Deutungsalternativen; unterhalb dieser Ebene existiert jedoch noch ein breites Feld in der Festlegung und Deutung syntaktischer und unterrichtspragmatischer Klassifikationsgesichtspunkte, die sich in der intersubjektiven Inkonsistenz der Beurteilung über alle Fächer hinweg bemerkbar machen (vgl. INGENKAMP, 1971). Dies gilt auch in gleichem Ausmaß für die Auswahl und Deutung der Klassifikationsgesichtspunkte für Antworten auf Testitems dann, wenn die sog. Interpretationsobjektivität zur Diskussion steht. Ihre Bewältigungsschwierigkeiten treten in methodischer Form als Problem der Trennschärfen-, Homogenitäts- bzw. Validitätsbestimmung von Testitems unter der Voraussetzung in Erscheinung, daß die Durchführungs- und Auswertungsobjektivität von Tests bereits abgesichert ist.

Als inhaltsanalytische Problem wird die Frage nach einer intersubjektiv konsistenten Klassifizierung von Testantworten nur stiefmütterlich zur Kenntnis genommen. In der testtheoretischen Diskussion gebraucht man dafür den Begriff der "Inhaltsvalidität" von Testitems. Da er keine methodische Formalisierung zuläßt, behandelt man ihn nur nebenbei, obgleich und weil er die

differenziertesten Ansprüche an die wissenschaftliche Begriffsbildung beinhaltet.

"Um die Inhaltsvalidität abschätzen zu können, muß das Ziel des Unterrichts in dem Gegenstand, in dem der Test angewandt werden soll, der Lernstoff, die relative Bedeutung der verschiedenen Unterrichtsmomente und so weiter explizit auf ihr G e d e i n e W e i s e (Sperrung vom Verfasser) bestimmt werden. Die Inhaltsvalidität kann nicht auf die gleiche Weise wie die prognostische oder gleichzeitige Validität in einem Validitätskoeffizienten ausgedrückt werden" (MAGNUSSON, 1969, S. 139).

Nach der Selbstverständlichkeit zu schließen, mit der man innerhalb der praktizierten Testkonstruktion und -anwendung das Problem der Inhaltsvalidität vernachlässigt, scheint sich die Bestimmung des Unterrichtszieles, der relativen Bedeutung der Unterrichtsmomente und so weiter bei Tests zu erübrigen, deren Validität durch einen Koeffizienten ausgedrückt werden kann. Auf der anderen Seite ist aber der Unterrichtsbezug solcher Tests notwendigerweise herzustellen, wenn die damit gewonnenen Urteile über Schüler prognostisch relevant sein sollen. Es stellt sich daher die Frage, wie die prognostische Relevanz von Urteilen über die Verhaltensprodukte von Schülern durch Tests sicherzustellen versucht wird und worin ihn Unterschied zur konventionellen Zensurengebung im wesentlichen besteht.

Es wurde vorher gesagt, daß die größere intersubjektive Inkonsistenz in der Zensurengebung z u n ä c h s t auf die Festlegung und den Deutungspielraum durch und für den individuellen Lehrer zurückzuführen ist. Dieser "Nachteil" der Zensurengebung gegenüber objektivierten Beurteilungsverfahren läßt sich jedoch nur solange ausweisen, wie Durchführungs- und Auswertungsobjektivität als Vergleichskriterien fungieren. Man kann ihn sofort beseitigen, wenn man die Durchführungsregeln von Prüfungen und die Auswertungsregeln für die Prüfungsergebnisse sukzessiv vereinheitlicht. In einer experimentellen Untersuchung haben MARTOG & MILES bereits 1936 nachgewiesen, daß die Beurteilung von Aufsätzen mit Hilfe eines detaillierten Schemas insgesamt zu übereinstimmenderen Ergebnissen bei verschiedenen

Prüfern führt als die Bewertung nach dem Eindruck. Durch die Standardisierung der Durchführungs- und Auswertungsregeln ergibt sich allerdings für die inhaltliche Deutung der übereinstimmenderen Ergebnisse keine veränderte Situation.

Dies zeigt sich sehr deutlich in einer kritischen Untersuchung der Urteilerobjektivität in der Bewertung von Aufsätzen, wo die Durchführungs- und Auswertungsbestimmungen durch ein Aufsatzbewertungsbeispiel vereinheitlicht wurden. Für die Formbewertung der Aufsätze ergab sich ein Objektivitätswert von 0.92, für die der Inhaltsbewertung 0.40 (RUCH & STODDARD, 1929). Unabhängig von dem Problem, wie sich allgemeinverbindliche Regeln für die Beurteilung von Inhalten überhaupt begründen lassen, beweisen die Objektivitätswerte, daß die Vereinheitlichung der Beurteilung sich in erster Linie auf die konsistentere Bewertung formaler Aspekte der Aufsatzgestaltung auswirkte, d.h. auf die Bewertung jener Aspekte, die durch reine Festlegung verbindlich gemacht werden können.

Der Umstand, daß unterschiedliche Auswahl und Deutung von formalen Gesichtspunkten zur Klassifikation von Schülerarbeiten eine wesentliche Ursache für die intersubjektive Inkonsistenz in der Gesamtbeurteilung sind, verdeutlicht sich auch bei der Bewertung von Mathematikarbeiten:

"Eine kleine Analyse kann ... zeigen, wie absurd die Annahme ist, eine Mathematikarbeit werde mit größerer Präzision beurteilt als eine sprachliche oder irgend eine andere Art von Prüfungsarbeit. Obwohl es stimmt, daß keine Meinungsunterschiede hinsichtlich der Richtigkeit der Lösung auftreten können, so gibt es doch zahllose Wege, auf denen die Lösung erreicht werden kann, wobei die Aufeinanderfolge der Schritte, der Gebrauch theoretischer Ansätze und Definitionen, die Genauigkeit von Zeichnungen, beispielhaft genannt werden können. Und dann geht es vor allem noch um den relativen Wert jeder speziellen Lösung und Definition in der Beurteilung der gesamten Arbeit. Ohne Zweifel ist die Komplexität der Faktoren nicht geringer in dieser Art von Prüfungen als in anderen" (STARCH & ELLIOT, 1913, S. 257f).

Die aufgezählten Faktoren sind ein Teilbereich der Bedingungen, die die intersubjektive Inkonsistenz in der Beurteilung erhöhen und die ausschließlich durch Konvention vereinheitlicht werden

könnten. Die Konventionalisierung der Klassifikation von Schülerarbeiten begegnet jedoch erheblichen Schwierigkeiten. Sie wird nicht nur durch die regionalen, strukturellen und organisatorischen Unterschiede zwischen den einzelnen Schulen in Bezug auf Einzugsbereich, sozioökonomische Struktur der Schülerpopulation, Lehrpläne, Lernmaterial usw. verhindert; solche Unterschiede wirken sich in verschiedenen Arten der Vermittlungstätigkeit und Unterrichtsgestaltung des einzelnen Lehrers aus und beeinflussen deren konkretes Klassifikationsverhalten. Hinsichtlich dieser Ursachen würde eine Konventionalisierung der Klassifikation die Beseitigung der genannten Schulunterschiede voraussetzen. Hier befindet sich auch die Zensurengebung gegenüber objektivierten Beurteilungsverfahren in keinem Nachteil, da die Testkonstrukteure mit dem gleichen Problem der Schulunterschiede konfrontiert sind.

Entscheidender ist die pragmatische Schwierigkeit, das Vorhaben der Konventionalisierung auf überregionaler Ebene zu organisieren und durchzuführen, auch dann, wenn an die Beseitigung der Schulunterschiede als notwendige Voraussetzung dafür nicht gedacht wird. Die Konventionalisierung wäre das Ergebnis der Konsensusbildung innerhalb der Gemeinschaft der beurteilenden Lehrer bzw. ihrer Vertreter. Man benötigt nicht viel Phantasie, um die Durchführbarkeit des Vorhabens unter gegebenen Verhältnissen als utopisch zu bezeichnen.

Bleibt die Beseitigung der Schulunterschiede als notwendige Voraussetzung für die Konventionalisierung der Beurteilung unberücksichtigt, dann scheint ihre Bewältigung durch die Konstruktion von standardisierten Beurteilungsverfahren im dagegen relativ kleinen Kreis der spezialisierten Forscher und Diagnostiker möglich zu sein. Der pragmatische Einwand verliert somit an Bedeutung. Gleichzeitig kommt aber in der Diagnostiker- und Forschergemeinschaft ein Konstruktionsverfahren für die Entwicklung von Beurteilungsmethoden zur Anwendung, das die Konventionalisierung der Gesichtspunkte für die Klassifikation von Verhaltensprodukten überflüssig erscheinen läßt. Es erweist

sich für die Herstellung der Durchführungs- und Auswertungsobjektivität des diagnostischen Prozesses als konstitutiv und erlaubt die Erklärung der geringeren intersubjektiven Konsistenz in der Zensurengebung.

Im Gegensatz zur Zensurengebung, wo die Vereinheitlichung in der Beurteilung der Schülerverhaltensprodukte zum Problem wird, geht es bei der Konstruktion von objektivierten Verfahren um die Vereinheitlichung der diagnostischen Anregungsbedingungen und um die Begründung ihrer Vergleichbarkeit. Im Prinzip ist mit der Vereinheitlichung der diagnostischen Anregungsbedingungen die Zielsetzung verbunden, die Reizlage für das zu untersuchende Individuum soweit zu standardisieren, daß die Verschiedenheit der Verhaltensprodukte einzelner Individuen nur noch auf die einzelnen Individuen selbst zurückgeführt werden kann. M.a.W.: Wenn konkrete Individuen auf die "g l e i c h e" Reizlage reagieren und verschiedene Reaktionsergebnisse aufweisen, dann kann das nur an den Unterschieden zwischen den konkreten Individuen liegen. Daraus läßt sich schlußfolgern: Wenn ausgezeichnete Mengen von Individuen, z.B. Schulklassen, Schüler verschiedener Schulgattungen usw., auf die gleiche Reizlage reagieren und verschiedene Reaktionsergebnisse aufweisen, dann können für die Unterschiede nur spezifische Eigenschaften der Mengen von Individuen verantwortlich gemacht werden.

Ich führe die Schlußfolgerung, die in der empiristisch orientierten Curriculum- und Unterrichtsforschung häufig gezogen wurde (vgl. SCHWARZ, 1973) und v.a. in varianzanalytischen Versuchsplanungen formalisiert wird (vgl. Ch. HOLZKAMP, 1973), an dieser Stelle auf, weil sich darin die Zwiespältigkeit der Interpretation von Urteilen über Schüler, erneut äußert. In Abhängigkeit von unterschiedlichen, methodischen Weiterverarbeitungen der mit dem Beurteilungsinstrument erhaltenen Rohdaten über die Verhaltensprodukte von Schülern verändert sich die Deutung des schülerspezifischen Kennwertes: Wird zum Beispiel eine bestimmte Population von Schülern über mehrere Regionen und Schulgattungen hinweg mit einem Testverfahren untersucht,

so interpretiert man das Ergebnis der rechnerischen Synthese der symbolisierten Aufgabenantworten als schülerspezifischen Ausprägungsgrad des erfaßten Merkmals. Alle Schüler erhalten somit je nach der Skalenqualität der Testresultate einen bestimmten Platz innerhalb der Gesamtpopulation. Berücksichtigt man dagegen nachträglich Region und Schulgattung der Schüler und stellt fest, daß beide Kennzeichen Modalwerte innerhalb der Gesamtverteilung der Testresultate erzeugen, dann lassen sich die Modalwerte nach einem Modell der veränderten rechnerischen Synthese der symbolisierten Rohdaten in die Bestimmung der schülerspezifischen Ausprägungsgrade miteinbeziehen. Dies führt zu einer veränderten Platzzuweisung für die Schüler in der Gesamtpopulation. Schüler aus Teilpopulationen mit niedrigen Modalwerten haben dann relativ hohe Ausprägungsgrade und umgekehrt. Manche Bestimmungen des Intelligenzquotienten durch standardisierte Verfahren gehen nach diesem Modus vor (AMTHAUSER 1953).

Sicherlich ist die unterschiedliche Platzzuweisung mit dem gleichen Beurteilungsverfahren nicht nur unter methodischem Gesichtspunkt zu diskutieren. Das unterschiedliche Vorgehen in der rechnerischen Synthese der symbolisierten Rohdaten ergibt sich ja erst aus der Aufteilung der Gesamtpopulation der Schüler in Teilpopulationen nach den Klassifikationsmerkmalen Region und Schulgattung. Auf der anderen Seite zeigt sich die Beurteilungsrelevanz beider Klassifikationsmerkmale nur an den vorliegenden Testresultaten, sodaß deren Konstituiertheit kritisch hinterfragt werden muß. Die durch das unterschiedliche Vorgehen bedingte Gegensätzlichkeit der Platzzuweisung wird dadurch zum Problem, daß mit Testverfahren der Anspruch verbunden wird, als Meßinstrumente für die Abbildung von Ausprägungsgraden dispositioneller Eigenschaften zu fungieren. Daraus ergeben sich zwei Arten von Schwierigkeiten:

1. Ein nicht auf empiristischem Weg eingeführter Dispositionsbegriff wird durch Testverfahren als metrischer, zumindest aber

als komparativer behandelt. Die Interpretationsschwierigkeiten von Testresultaten sind nicht zuletzt auf die Vermischung beider Begriffsarten bzw. auf die Vernachlässigung der Bedingungen für ihre Einführung in die Wissenschaftssprache zurückzuführen. Die Rekonstruktion dieser Schwierigkeiten kann hier nicht erfolgen, da sie die eingehende Behandlung der Einführung von metrischen Begriffen in die Wissenschaftssprache voraussetzt (vgl. STEGMÜLLER, 1970, Band II, Studienausgabe Teil A).

2. Ein Meßinstrument als Mittel der Abbildung von quantitativen Eigenschaften von Gegenständen zeichnet sich gerade dadurch aus, daß es unabhängig von der Menge der Gegenstände gebraucht werden kann, über deren Vergleich es entwickelt wurde und auf die es angewandt werden soll. Die Abhängigkeit des Lehrerurteils vom k o n k r e t e n Schüler in seiner Gesamtpersönlichkeit bzw. von der Gruppe der Schüler, d.h. von Bedingungen, die für die Qualifikationsbeurteilung keine Rolle spielen sollten, ist einer der wichtigsten Gründe für die intersubjektive Inkonsistenz in der Deutung der Beurteilungskategorien durch die verschiedenen Lehrer.

Mit objektivierten Beurteilungsverfahren soll die Schwierigkeit dadurch bewältigt werden, daß der urteilende Lehrer aus dem Beurteilungsprozeß weitgehend ausgeschaltet wird. Nun diskriminieren aber auch e n t s u b j e k t i v i e r t e Beurteilungsverfahren zwischen Schülern in Bezug auf Dimensionen, für die sie nicht entwickelt wurden, was z.B. in allen sprachgebundenen Tests überaus deutlich hervortritt. Der "Vorteil" ent-subjektivierter Verfahren besteht also in der Möglichkeit der methodischen Objektivierung der genannten Abhängigkeit.

Bei der Zensurengebung bereitet die Objektivierung dagegen außerordentliche Schwierigkeiten. Durch oben erwähnte Verfahren der Modalwertberücksichtigung bei der Bestimmung der Ausprägungsgrade dispositioneller Eigenschaften läßt sich die Abhängigkeit der Beurteilung von der untersuchten Population zwar auch nicht ausschalten, aber methodisch formalisiert darstellen.

2.4.5.2 Meßinstrumentelle Bedingungen objektivierter Schülerbeurteilung

Die bisherige Erörterung der Vorzüge objektivierter Schülerbeurteilung begründet die Vermutung, die Funktion standardisierter Beurteilungsverfahren beschränke sich in der methodisch formalisierten Darstellung bekannter und traditioneller Schwierigkeiten innerhalb der praktizierten Diagnostik in der Schule. Es ist deshalb zu untersuchen, ob entsubjektivierte Beurteilungsverfahren den Namen "Meßinstrument" zurecht tragen. Dazu ist von einer Unterscheidung zwischen zwei Funktionen von Meßinstrumenten auszugehen, die in der empiristisch orientierten Sozialwissenschaft kaum Beachtung fand und an die FISCHER (1974) wieder erinnert hat.

Meßinstrumente können im Dienst der Metrisierung stehen und zur eigentlichen Messung herangezogen werden:

"Die Einführung eines quantitativen Begriffes für einen Bereich von Objekten wird ... als Metrisierung bezeichnet. Von der Metrisierung ist die Messung zu unterscheiden, worunter man den empirischen Prozeß der Bestimmung eines Größenwertes versteht. Messungen können erst vorgenommen werden, wenn bereits eine Metrisierung vorliegt. Leider bezeichnen viele Autoren das, was wir Metrisierung nennen, als Messung. Es ist nicht verwunderlich, daß bei einer derartigen irreführenden Terminologie leicht Begriffsverwirrungen entstanden sind" (STEGMÜLLER, 1970, Band II, Studienausgabe Teil A, S. 46).

Es wäre empirisch sinnlos von der Messung einer Größe zu reden, d.h., man könnte keinen Anspruch erheben, sich in einer empirischen bedeutungshaltigen Sprache über Größenwerte zu unterhalten, wenn die Größe nicht als metrischer Begriff in die Wissenschaftssprache eingeführt ist. Allgemein werden mit der Metrisierung eines noch nicht quantifizierten Begriffes zwei Bestimmungen unternommen:

1. Ist ein noch nicht quantifizierter Begriff F - z.B. "größer als", "härter als", "schwerer als" - auf einer Menge von Objekten M definiert, d.h., ist man sich einig, was unter

"größer als", "schwerer als" und "härter als" zu verstehen ist, dann gilt es eine Funktion $m(x)$ anzugeben, die die folgende Bedingung erfüllt:

Sie muß auf M definiert sein und M in die Menge der reellen Zahlen abbilden;

2. Es gilt ferner einen auf der Menge der reellen Zahlen definierten Begriff f anzugeben, mit dem sich die Bestimmungen, die F für die Objekte x der Menge M trifft, für die diesen Objekten zugeordneten Zahlen $m(x)$ darstellen lassen (vgl. v. KUTSCHERA, 1972, S. 26).

Für eine knappe Veranschaulichung sei v. KUTSCHERAs Beispiel angeführt:

"So kann man z.B. dem komparativen Begriff der Härte $H(x,y)$ - x ist höchstens so hart wie y - einen metrischen Begriff des Härtegrades $h(x)$ zuordnen, für den gilt: $H(x,y) \equiv h(x) \leq h(y)$, d.h.: das Mineral x ist höchstens so hart wie das Mineral y genau dann wenn der Härtegrad von x , d.h. die Zahl $h(x)$, kleiner oder gleich dem Härtegrad von y , $h(y)$ ist. (Hier ist also F der Begriff H und f die Beziehung $\leq$)" (v. KUTSCHERA, 1972, S. 26).

Diese allgemeinen Bestimmungen für die Metrisierung von Begriffen repräsentieren jedoch nur die Darstellungsform ihrer Einführung in die Wissenschaftssprache ¹³⁾. Zur Realisierung der Einführung benötigt man Meßinstrumente, die die geforderte formale Beziehung zwischen den beiden Mengen der Objekte und der reellen Zahlen zu vermitteln in der Lage sind. Für die Herstellung dieser Beziehung sind eine Reihe von Regeln zu befolgen, wovon eine bestimmte in Bezug auf die Funktion der Meßinstrumente von besonderer Bedeutung ist. Es handelt sich um das Prinzip der Additivität, das sich in der allgemeinen Form darstellen läßt:

-
- 13) Auf eine Behandlung der komplexeren Einführung von metrischen Begriffen als abgeleitete Größen und der Unterscheidung von extensionalen und intensionalen Größen muß hier verzichtet werden. Im Zusammenhang mit der Frage der Interpretierbarkeit von diagnostischen "Meßresultaten" interessiert nur die oben skizzierte fundamentale Metrisierung, zumal die abgeleitete auf sie zurückführbar ist.

$$f(a \circ b) = f(a) + f(b).$$

Die Formel bedeutet, daß der Summe der numerischen Werte, die den Objekten x der Menge M durch die Funktion f zugeordnet werden, nämlich die Werte $f(a)$ und $f(b)$, auf der linken Seite der Gleichung ein Wert $f(a \circ b)$ entspricht; dieser Wert ist nur dann eindeutig bestimmt, wenn die Objekte a und b in spezifischer Weise miteinander kombiniert werden. Die Kombinationsoperation ist hier durch das Symbol $\circ$ charakterisiert. Für ihren Vollzug ist ein Meßinstrument zu finden, das die Entsprechung von beiden Seiten der Gleichung für alle Objekte der Menge M bestätigt. Neben konventionellen Gesichtspunkten bei der Auswahl des geeigneten Meßinstruments spielt vor allem die Tatsache eine Rolle, daß es empirische Relationen zwischen den zu messenden Objekten indiziert und seine Funktionsweise durch empirische Regelmäßigkeiten bestimmt ist. So ist etwa die Gleichgewichtsrelation und die Relation "schwerer als" eine empirische Relation, über deren Vorliegen im konkreten Fall durch eine empirische Untersuchung entschieden werden muß. Die Untersuchung erfordert ein Meßinstrument, daß die empirischen Relationen zwischen den Objekten in naturgesetzlicher Weise indiziert d.h., bei wiederholter Anwendung des Meßinstruments auf die gleichen Objekte muß man die gleichen Ergebnisse erhalten. Da die Brauchbarkeit des Meßinstruments nicht als allen Objekten zu überprüfen ist, gehen in ihre Begründung für die Verwendbarkeit zur Einführung metrischer Begriffe in die Wissenschaftssprache empirische Hypothesen ein, die im besten Fall mehr oder weniger gut bestätigt sein können.

"Nehmen wir an, es liege ein komparativer Begriff $x \leq y$ vor und es sei aufgrund von Naturgesetzen bekannt, daß der Zeiger des Meßinstruments bei einer Messung von x höchstens soweit ausschlägt wie bei einer Messung von y genau dann, wenn $x \leq y$... Man mißt dann ... mit dem Meßinstrument keine von ihm unabhängige gegebene metrische Größe ..., sondern man stellt nur fest, ob die Relation $x \leq y$ vorliegt oder nicht (v. KUTSCHERA, 1972, S. 30f).

Das Meßinstrument fungiert also als Indikator für das Vorliegen von empirischen Relationen zwischen Objekten. Erst durch die

Bestätigung der Existenz solcher Relationen wird die Voraussetzung dafür geschaffen, dieses oder ein anderes Meßinstrument durch hinzukommende Skalierung und Eichung für die Bestimmung von Größenwerten, d.h. für die Messung im eigentlichen Sinn, einzusetzen. Daß Naturwissenschaftler indizierende Meßinstrumente zuverlässig funktionierend bereits vorfinden und nicht in jedem Fall erst entwickeln müssen, ist ein Problem, das durch die Untersuchung der Geschichte der Naturwissenschaften geklärt werden müßte. Sie würde zeigen, wie der empiristische Bestätigungsbegriff mit vorwissenschaftlichen Erfahrungsregelmäßigkeiten im Umgang mit bereits früher existierenden Meßinstrumenten in Zusammenhang zu bringen ist.

Im Hinblick auf die Durchsetzung des Additivitätsprinzips bei der Metrisierung von Begriffen steht hier die Indikatorfunktion von Meßinstrumenten im Vordergrund des Interesses.

Mit Hilfe des Meßinstrumentes müssen sich die Objekte so kombinieren lassen, daß das Additivitätsprinzip - neben anderen Bedingungen - erfüllt ist. Am Beispiel des Meßinstrumentes "Balkenwaage" zur Einführung des metrischen Begriffes des Gewichtes kann die Kombination so erfolgen, daß man zwei als gleich schwer ermittelte Objekte a und b in eine Schale der Balkenwaage legt. Man stellt fest, daß die beiden Objekte sich mit einem dritten Objekt in der anderen Schale im Gleichgewicht befinden. Ordnet man nun den beiden Objekten a und b jeweils die Ziffern 1 durch eine Funktion f zu, dann läßt sich der Kombination (a o b) die Ziffer 2 zuordnen, sodaß nachträglich das Prinzip der Additivität erfüllt ist.

Wesentlich dabei ist, daß die Ziffer 2 der Kombination (a o b) nicht aufgrund der Addition der beiden Ziffern 1 zugeordnet wurde, sondern aufgrund des empirischen Ergebnisses mittels Meßinstrument. Dies setzt die Gegebenheit des empirischen Sachverhaltes voraus, daß das dritte Objekt durch die Kombination zweier gleichschwerer Objekte in der erfolgten Weise aufgewogen werden kann und daß die Balkenwaage keine Mängel aufweist. Bei der Einführung des metrischen Begriffes der Länge ist z.B.

der empirische Sachverhalt vorauszusetzen, daß das Meßinstrument "Stab" in seiner Längenausdehnung relativ konstant bleibt, während für die Metrisierung der Zeit ein relativ stabiler Fendelvorgang erforderlich ist.

Erst wenn für die Realisierung des Additivitätsprinzips ein Meßinstrument vorhanden ist, mit dem sich die verschiedenen Objekte der Menge M in dieser oder einer entsprechend ähnlichen Weise kombinieren lassen, kann man zur Konstruktion des Meßinstrumentes im engeren Sinn übergehen. Dazu ist allerdings die Erfüllung von weiteren Bedingungen und Regeln erforderlich wie die Festlegung eines Einheitswertes und Standardobjektes, die Überprüfung der Eindeutigkeit der Metrisierung, d.h. die Überprüfung der Invarianz der Metrisierungsfunktion gegenüber Transformationen, sowie die Überprüfung der Erfüllung bestimmter Maßprinzipien. Bei Gegebenheit dieser Voraussetzungen kann man dann an die Konstruktion eines pragmatisch brauchbaren Instrumentes zur Messung von Größenwerten übergehen und die zu verwendende Skala und Eichung konventionalisieren. Hier stellt sich dann die Frage nach der empirischen Bedeutung der Meßwerte nicht mehr, da sie bereits bei der Metrisierung des Begriffes beantwortet wurde.

Will man in der empiristisch orientierten Diagnostik mit metrischen Begriffen umgehen, um die Aufteilung von Schülern auf weiterführende Qualifizierungsalternativen durch den Gebrauch von quantifizierbaren Lernvoraussetzungen zu rechtfertigen, dann ist dieses Vorgehen nach den Kriterien für die Einführung von metrischen Begriffen in die Wissenschaftssprache zu bemessen. Eines der entscheidendsten Bemessungskriterien ist die Verfügung über eine Operation zur Kombination von Objekten mittels eines Meßinstrumentes, das die empirischen Relationen zwischen den Objekten zu indizieren vermag. Das Verfügen über eine derartige Operation wäre ein konstitutiver Bestandteil innerhalb des Regelkomplexes für die Einführung eines metrischen Eigenschaftsbegriffes. Da von einem Test behauptet wird, ein Meßinstrument zu sein, muß gefragt werden, ob er zur Metrisierung von M -Eigenschaftsbegriffen beiträgt und ob er auf dieser

Basis zu einem Meßinstrument im engeren Sinn, also zu einem Instrument zur Bestimmung von Größenwerten - den Ausprägungsgraden dispositioneller Eigenschaften -, entwickelt werden kann.

2.4.5.3 Das Problem der empirischen Deutung von "Meßwerten" in der Schülerbeurteilung

Allgemein besteht ein nach den Kriterien der klassischen Testtheorie konstruierter Test aus einer Menge von Items, von denen angenommen wird, daß sie die zu messende Dimension repräsentieren und unterschiedliche Ausprägungsgrade dieser Dimension indizieren. Gewöhnlich wird aber bei der Beurteilung der untersuchten Individuen die in den einzelnen Items enthaltene Information nicht ausgenutzt, da die Antworten auf die einzelnen Items die untersuchte Population jeweils nur in zwei Klassen einteilen: in die Klasse der Richtiglöser und die der Falschlöser. In der Form nach gehandhabter, quantitativer Begriff für die Beurteilung kommt erst dadurch zustande, indem die durch Ziffern (1 und 0) symbolisierten Antworten auf die Items zu einem Summenwert zusammengefaßt werden.

Dieser Summenwert wird aber bereits als Bestimmung des Größtenwertes einer individualspezifischen Schülereigenschaft interpretiert, d.h., der Test fungiert bereits als Instrument zur Messung im engeren Sinn. Inwiefern und an welcher Stelle im Vollzug der Testkonstruktion wurde dabei ein metrischer Begriff eingeführt ?

Das Vorgehen in der Summenwertbildung hat auf den ersten Blick eine gewisse, formale Ähnlichkeit mit der Realisierung des Additivitätsprinzips. Den verschiedenen Objektiven "Lösungsprodukte" wird entweder die Ziffer 1 oder die Ziffer 0 zugeordnet; die Summierung der numerischen Werte der Lösungsprodukte ergibt einen neuen Wert, dessen Konstituiertheit sich aber

nicht in der arithmetischen Operation der Summierung erschöpfen dürfte, wenn das Vorgehen die Einführung eines metrischen Begriffes begründen soll. Vorausgesetzt, es liegen mehrere Verhaltensprodukte $v_1, \dots, v_n$ als Resultat der Auseinandersetzung eines Schülers mit verschiedenen Items $i_1, \dots, i_n$ vor, dann müßten sich die verschiedenen Verhaltensprodukte - formal betrachtet - folgendermaßen kombinieren lassen, wenn durch sie ein metrischer Begriff operationalisiert werden soll:

$$f(v_1 \circ v_2 \circ \dots \circ v_n) = f(v_1) + f(v_2) + \dots + f(v_n)$$

Um die Gültigkeit der formalen Beziehung zu begründen, sind zwei Probleme zu bewältigen:

1. Aufgrund welcher Eigenschaften der Verhaltensprodukte wird einem Teil davon der Wert 1 zugeordnet und einem anderen Teil der Wert 0 ?
M.a.W.: Aufgrund welcher Eigenschaften sind die Verhaltensprodukte innerhalb der beiden Klassen gleich ?
2. Welche empirische Operation "o" entspricht der arithmetischen Operation der Addition und auf welchen empirischen Befunden bzw. empirischen Gesetzmäßigkeiten baut die Kombination der verschiedenen Verhaltensprodukte durch "o" auf ?

Da auch die Zuordnung der Werte 1 und 0 unter Bezug auf empirische Gesetzmäßigkeiten erfolgen muß, die für die Kombinationsoperation konstitutiv ist, lassen sich beide Fragen nicht unabhängig voneinander beantworten.

Um die verschiedenen Verhaltensprodukte in zwei Klassen aufteilen zu können, deren Elemente gleiche empirische Eigenschaften haben, bedarf es eines Meßinstrumentes, das die empirischen Relationen zwischen den Verhaltensprodukten abbilden kann. Von einem Test wird behauptet, daß er die Funktion eines Meßinstrumentes erfüllt. Streng genommen müßte deshalb der Test als zusammenhängende Gesamtheit verschiedener Items indizieren, daß

die verschiedenen Verhaltensprodukte eines Schülers in zwei Klassen aufteilbar sind. Dies ist aber nicht möglich, da er ein breites Spektrum verschiedener Summenwerte erzeugt und die Berechnung der Summenwerte die zu indizierende Klassifikation der Verhaltensprodukte voraussetzt. In Analogie zur Funktion des Meßinstrumentes "Balkenwaage" versagt damit der Test als zusammenhängende Gesamtheit aller Items, die Gleichheit bzw. Verschiedenheit der Objekte "Verhaltensprodukte" zu indizieren.

Im Unterschied dazu zeigt er an, daß es Objekte gibt - die zu untersuchenden Individuen -, die die zu klassifizierenden Objekte produzieren. Beim Produktionsprozeß der Objekte handelt es sich aber um den zweiten Gegenstandsbereich der Diagnostik, der einer unmittelbaren Metrisierung nicht zugänglich ist und dessen Quantifizierbarkeit eben durch die Metrisierung des ersten Gegenstandsbereiches rekonstruiert werden soll. Zudem kann die Funktion des Tests, zwischen verschiedenen Individuen zu diskriminieren, erst dann begründet werden, wenn der allgemeine Zusammenhang zwischen den diagnostischen Anregebungsbedingungen "Testitems" und den Verhaltensprodukten geklärt ist. Es ist daher erforderlich, die einzelnen Bestandteile eines Tests zu untersuchen, um zu ermitteln, auf welche Weise die Vergleichbarkeit der Verhaltensprodukte sichergestellt wird. Die Klärung des allgemeinen Zusammenhanges zwischen den einzelnen diagnostischen Anregebungsbedingungen im Test und den verschiedenen Verhaltensprodukten ist gleichzusetzen mit der Frage nach dem, was für alle untersuchten Individuen in gleicher Weise gilt. Infolgedessen kann man zu analytischen Zwecken von dem Schüler ausgehen und muß untersuchen, ob sich seine Verhaltensprodukte mittels dem Meßinstrument in zwei Klassen aufteilen lassen.

Hier ist nun einerseits festzuhalten, daß die Klassifikation der Verhaltensprodukte vor der Entwicklung und Konstruktion des Meßinstrumentes ohne die Kenntnis von empiristisch gewonnenen Befunden bereits festgelegt ist. So legitimiert sich ein Matho-

matiktest, bestehend aus einer Reihe von Dreisatz-Textaufgaben, nicht dadurch, daß die einzelnen **m a n i f e s t e n** Lösungsprodukte nachträglich untereinander Ähnlichkeiten aufweisen; die Ähnlichkeiten sind vielmehr vorbestimmt durch die Auswahl von Rechenoperationen, deren Erwerb durch den Schüler wiederum durch Festlegung im Ausbildungssystem vorgesehen ist und deren Beherrschung durch entsprechende Textaufgaben überprüft werden soll. Auch die Bewertung verschiedener Lösungswege zum Verhaltensprodukt richtet sich letztlich nach disziplinstrukturellen Gesichtspunkten und damit nach analytischen Bestimmungen des Prüfungsgegenstandes. Gleiches wurde bei der Behandlung der Klassifikationskategorien für die Aufsatzbeurteilung schon aufgezeigt. In dieser Hinsicht besteht auch kein Unterschied zur Klassifikation von Antworten auf Items aus Intelligenz- und anderen Tests. Die Klassifikationskategorien beinhalten Merkmale, die unabhängig vom konkreten Problemlösungsprozeß in der diagnostischen Situation festgelegt werden und auf die manifesten Verhaltensprodukte nachträglich Anwendung finden.

Obgleich die Klassifikationskategorien von Verhaltensprodukten wie z.B. "richtig - falsch", "brauchbar - unbrauchbar" oder "erwünscht - unerwünscht" insofern nicht erfahrungsunabhängig sind, als sie zwischen den verschiedenen Schülern diskriminieren, ist ihre Bedeutung nur noch analytisch zu bestimmen. Das heißt aber, daß für die Entscheidung, ob ein manifestes Verhaltensprodukt unter eine der beiden jeweils dichotomen Klassifikationsalternativen zu subsumieren ist, kein empirisch begründetes Meßinstrument zur Verfügung stehen kann. Die Zuordnung der Ziffern 1 und 0 legitimiert sich ausschließlich durch die disziplinstrukturellen Klassifikationsgesichtspunkte, die sich formal ebenso unterscheiden wie die Ziffern 1 und 0. Sie müssen für den jeweiligen Prüfungsgegenstand gesondert festgelegt werden und sind dann zu verändern - und dies belegt ihre empirische Bedeutungslosigkeit im empiristischen Sinn aus der Perspektive des Selektionscharakters von Beurteilungen -, wenn

ihr Beitrag zur Diskriminierung zwischen den Schülern abnimmt, z.B., wenn zuviele Schüler Richtiglösungen bzw. Falschlösungen produzieren.

Infolgedessen fehlt der Addition der zugeordneten Ziffern auf der rechten Seite der Gleichung für das Additivitätsprinzip die notwendige Entsprechung auf der linken Seite. Für den Operator "o" läßt sich keine Interpretation angeben, die sich auf einen empirischen Befund, eine empirische Hypothese oder eine empirische Gesetzmäßigkeit zurückführen ließe. Der Operator repräsentiert hier die definitorische Festlegung durch den Diagnostiker oder Beurteiler, wonach zwei oder mehrere Verhaltensprodukte deshalb als gleich zu klassifizieren sind, weil sie die gleichen, disziplinstrukturell geforderten Merkmale haben. Dadurch übernimmt der Beurteiler, Diagnostiker oder Forscher die Funktion des Meßinstrumentes in der gleichen Weise wie bei der Zensurengebung. Im Gegensatz zur objektivierten Beurteilung fehlen aber bei der Zensurengebung die standardisierten, verhaltensprovozierenden Anregebungsbedingungen in Form der einzelnen Testitems.

Zum anderen besteht zwischen Zensurengebung und objektivierten Beurteilungsverfahren ein über die Standardisierung hinausgehender Unterschied. Um die Funktion des Diagnostikers als Meßinstrument und als Operator für die Kombination der verschiedenen Verhaltensprodukte im konkreten Beurteilungsprozeß soweit wie möglich einzuschränken, müssen die einzelnen Testitems als verhaltensprovozierende, diagnostische Anregebungsbedingungen maximal homogenisiert werden. Jedes Item sollte also die Produktion von solchem manifesten Verhalten zulassen, für dessen Klassifikation die festgelegten Beurteilungskategorien schon bereit stehen.

Bei der Zensurierung von Aufsätzen ist diese Forderung im extremen Ausmaß nicht erfüllt. Es gibt in der Regel nur eine diagnostische Anregebungsbedingung, die Aufforderung, einen Aufsatz mit diesem oder jenem Thema zu schreiben, aber mehrere, voneinander unabhängig angewandte Regeln für die Aufsatzbeurteilung.

Da diese einzige diagnostische Anregebungsbedingung keine Hinweis enthält, welche Merkmale in welcher Häufigkeit der geschriebene Aufsatz aufweisen soll, fällt dem urteilenden Lehrer die gesamte Last der Merkmalsfestlegung und der Kombination der Bestandteile eines Aufsatzes zu. Bei der Zensurierung anderer Prüfungsgegenstände wie etwa von Mathematikarbeiten, von Diktaten, von Übersetzungen eines Textes in eine Fremdsprache oder Arbeiten aus naturwissenschaftlichen Fächern ist der Komplex diagnostischer Anregebungsbedingungen bisweilen differenzierter aufgebaut und nach Teilaufgaben untergliedert, die in sich schon Gesichtspunkte für die Beurteilung der potentiellen Verhaltensprodukte beinhalten. Dennoch hängt die Anwendung der Klassifikationsregeln vom produzierten, manifesten Verhalten ab.

Durch das Festlegen der Klassifikationsregeln für manifeste Verhaltensprodukte und durch die Tatsache, daß die Klassifikationsmerkmale als dichotome Pole einer Dimension logisch voneinander abhängig sind, müssen bei der Entwicklung von Tests die einzelnen Items so konstruiert werden, daß die richtige, erwünschte oder brauchbare Lösungsalternative vom Schüler grundsätzlich produziert werden kann. Die Items müssen also ein Minimum an Information enthalten, deren Abbau bzw. Transformation durch den Schüler zu der vorgesehenen positiven Lösungsalternative führen könnte. Ist diese Voraussetzung für alle Items gegeben und kann man ferner voraussetzen, daß die positiven Lösungsalternativen in Bezug auf den Prüfungsgegenstand gleich oder zumindest ähnlich sind - z.B. sich auf das gleiche ausgewählte Stoffgebiet beziehen oder die gleichen Umgangsformen mit dem Stoffgebiet prüfen wie Lösen von Gleichungen, Anfertigen von originaltreuen Zeichnungen, fehlerloses und treffendes Beschreiben eines Vorganges usw. -, dann scheint sich die Anzahl der tatsächlich erreichten, positiven Lösungsalternativen als Begründung für die Einführung eines metrisierbaren Begriffes verwenden zu lassen. Da aber die verschiedenen Items wegen der unterstellten Voraussetzungen die notwendige und hinreichende Bedingung für die Möglichkeit des Auftretens der in-

nen entsprechenden, positiven Lösungsalternativen darstellen, können auch sie nicht als Meßinstrumente für die Gleichheit oder Verschiedenheit der Verhaltensprodukte fungieren. Jedes Item läßt eine positive und viele negative Lösungsmöglichkeiten zu. Wenn die positiven Lösungsmöglichkeiten der verschiedenen Items eines Tests zum Zweck der vorgesehenen Addition als gleiche beurteilt werden sollen - für die negativen Lösungsalternativen gilt dies wegen ihrer logischen Abhängigkeit von den positiven ohnehin -, dann muß die unterstellte Voraussetzung als gesichert gelten, daß die Items die notwendige und hinreichende Bedingung für die Möglichkeit des Auftretens der ihnen entsprechenden positiven Lösungsalternativen in der Tat bilden. Sollte dies eine sinnvolle und belegbare Voraussetzung sein, so besteht das Forschungsproblem nicht mehr in der Entwicklung eines Meßinstruments für die Gleichheit der Verhaltensprodukte; letzteres ist ja durch den Katalog der Qualifikationsanforderungen per definitionem festgelegt. Zu entwickeln wäre vielmehr, wenn sich die Voraussetzung als richtig erwies, ein Meßinstrument für die Gleichheit der verhaltensprovozierenden Items.

Damit steht aber nicht mehr die Metrisierbarkeit der Objekte "Verhaltensprodukte" und der Objekte "Schüler" zur Diskussion, sondern die der Objekte "Items" als verhaltensprovozierende, diagnostische Anregungsbedingungen.

Aus diesem Grund ist die Interpretation des Additivitätsprinzips in der vorher erwähnten Weise zu verändern, indem an die Stelle der Verhaltensprodukte die verschiedenen Items treten:

$$f(i_1 \circ i_2 \circ , \dots , \circ i_n) = f(i_1) + f(i_2) + , \dots , + f(i_n)$$

Die Realisierung des Additivitätsprinzips in dieser Form ist nur dann für die Metrisierung der Items möglich, wenn die Beziehung zwischen den verschiedenen Items und den jeweiligen positiven Lösungsalternativen in ihrer empirischen Bedeutung

aufgeklärt werden kann; d.h.: Voraussetzung für die Realisierung sind Kenntnisse über die Spezifität der Auseinandersetzung mit den diagnostischen Anregebungsbedingungen all jener untersuchten Individuen, die positive Lösungsalternativen produzieren. Für alle anderen kann die Realisierung solange nicht erfolgen, wie die negativen Lösungsalternativen ausschließlich als kontradiktorischer Gegensatz zu den positiven bestimmt, also nur als logisch abhängig definiert sind. Die Realisierung des Additivitätsprinzips bestünde dann darin, mit Hilfe eines Meßinstrumentes die verschiedenen Items miteinander zu vergleichen, die gleichen auszuwählen und ihnen Zahlen in der Weise zuzuordnen, daß dem Summenwert dieser Zahlen ein Wert entspricht, der den kombinierten Items zugeordnet wurde.

Geht man von der unterstellten Voraussetzung aus, daß sich für alle Beziehungen zwischen den gleichen Items und den entsprechenden positiven Lösungsalternativen eine empirische Begründung angeben läßt, daß also eine beziehungskonstitutive, dispositionelle Eigenschaft im Schüler bekannt ist, dann könnte man die Anzahl der produzierten, positiven Lösungsalternativen als Ausprägungsgrad der dispositionellen Eigenschaft zwar festlegen aber nicht begründen. Dies wäre allerdings der Form nach bereits eine Messung im engeren Sinn und würde nur für die Gesamtheit aller Schüler gelten, die in die Realisierung des Additivitätsprinzips nicht eingehen, weil durch ihr Lösungsverhalten die Gleichheit der Items nicht bestätigt werden konnte.

Der Nachweis der Gleichheit von Items läßt sich also nur auf zwei Wegen erreichen:

1. Man wählt jene Items aus, die von allen untersuchten Individuen mit der positiven Alternative beantwortet wurden.
2. Man beschränkt die Gültigkeit der Gleichheit auf jene Individuen, die mit der positiven Alternative antworten.

Das erste Verfahren entspricht der Itemcharakterisierung bei der Entwicklung von kriterienorientierten Beurteilungsverfahren (vgl. SPITTELMAYER, 1973), das zweite der Itemcharakterisierung

bei normorientierten Tests. In beiden Fällen ist die Charakterisierung von der Population der untersuchten Individuen abhängig, zu deren Beurteilung die charakterisierten Items dienen sollen. Für die Zirkularität des Vorgehens ist die Tatsache verantwortlich zu machen, daß man in der Diagnostik vor der Entwicklung eines Beurteilungsinstrumentes über keinerlei empirisch bedeutungshaltige Kenntnis von dispositionellen Eigenschaften verfügt und - wie vorher ausgeführt wurde - nicht in der Lage ist, Dispositionsbegriffe auf empiristischem Weg in die Wissenschaftssprache einzuführen. Die praktizierte Quantifizierung von dispositionellen Eigenschaften setzt somit voraus, was erst nachzuweisen wäre. Da dabei einunddasselbe Vorgehen zur Anwendung kommt und die Qualität des Vorgehens selbst unter Beweis gestellt werden soll, überträgt man dem untersuchten Individuum implizit die gesamte Bestätigungsfunktion.

Daraus ergibt sich eine widersprüchliche Konsequenz für die Interpretation des Operators "o", der ja eine auf empirischen Gesetzmäßigkeiten aufbauende Methode zur Kombination der Einzelitems repräsentieren sollte:

Vom Anspruch der Testkonstrukteure her, müßte der Test ein Instrument sein, das in Abhängigkeit vom untersuchten Individuum unterschiedlich stark ausschlägt. Der Zeigerausschlag konkretisiert sich bei einem Test in den produzierten, positiven Lösungsalternativen. Da allen positiven Lösungsalternativen die Zahl 1 zugeordnet wird und sie mithin als gleichwertig betrachtet werden, sind die Items untereinander ebenfalls als gleichwertig anzusehen. Da gleichzeitig die Anzahl der produzierten, positiven Lösungsalternativen als Ausprägungsgrad der zu messenden dispositionellen Eigenschaft interpretiert wird, bedeutet dies, daß durch die Anzahl der gleichwertigen, dargebotenen Items der Zeigerausschlag des Meßinstrumentes festgelegt wird. Je mehr Items von einem untersuchten Individuum bearbeitet u n d mit der positiven Lösungsalternative beantwortet werden, desto höher ist der Zeigerausschlag des Tests. Damit

also der Test als Kombination vieler, gleichwertiger Items betrachtet werden kann, bedarf es einer adäquaten Beantwortung der Items durch das untersuchte Individuum. Eine vom untersuchten Individuum unabhängige, empirische Methode zur Kombination der Items im Sinne der Realisierung des Additivitätsprinzipes existiert nicht.

Somit ist das untersuchte Individuum die empirisch objektive Basis für die Kombination der Items zu einem Gesamttest, übernimmt dadurch die Funktion des Operators "o" und stellt deshalb das Meßinstrument für den Vergleich und die Kombination der verschiedenen Items dar. Man kann infolgedessen von einem Test, der eine dem untersuchten Individuum unterstellte, dispositionelle Eigenschaft zu messen vorgibt, dann sprechen, wenn das untersuchte Individuum die Anzahl der dargebotenen Items bearbeitet und mit den festgelegten, positiven Lösungsalternativen beantwortet. Damit würde aber die Gültigkeit des Additivitätsprinzipes als metrische Beziehung zwischen Items durch den Operator "untersuchtes Individuum" bestätigt werden, obwohl das untersuchte Individuum selbst Objekt der Messung sein soll und der Test als Meßinstrument metrisierte Eigenschaften von ihm abbilden müßte.

In dieser Art der Realisierung des Additivitätsprinzipes ist das zu untersuchende und untersuchte Individuum das einzige empirische Moment für die Metrisierung der verschiedenen Items. Es ist zugleich Gegenstand und Methode der Metrisierung, indem es die vom Forscher bzw. Diagnostiker festgelegte Klassifizierung seiner Verhaltensprodukte als Resultat der Auseinandersetzung mit den verschiedenen diagnostischen Anregungsbedingungen bestätigt. Grundlage für die Beurteilung verschiedener untersuchter Individuen kann deshalb nur das unterschiedliche Ausmaß sein, mit dem die festgelegte Klassifizierung jeweils bestätigt wird. Das heißt: Die Quantifizierung dispositioneller Eigenschaften erschöpft sich in der Zählung positiver Verhaltensprodukte. Da die Anzahl der Verhaltensprodukte von der Anzahl der Items abhängt, bestimmt sich der Zeigerausschlag ausschließ-

lich nach einer notwendigen Länge des Tests, wobei die Notwendigkeit wiederum mit der geforderten Häufigkeit von Unterscheidungen zwischen Individuen zusammenhängt.

Zusammenfassung

Konkrete Schüler, die über die vorausgesetzte empirische Eigenschaft verfügen, die Klassifikation ihrer Verhaltensprodukte durch den Forscher bzw. Diagnostiker nach dem geschilderten Verfahren zu bestätigen, konstituieren objektiv den Zusammenhang zwischen den verschiedenen Items. Sie konstituieren eine wiederholbare, invariante Beziehung zwischen den verschiedenen Items und den jeweiligen Antworten darauf und bestätigen als Teilpopulation aller Schüler die Hypothese des Forschers, daß die Anwendbarkeit seiner Klassifikation auf eine Reihe von Verhaltensprodukten möglich ist, wenn die Produktion dieses wiederholbaren Verhaltens durch äquivalente Items provoziert wird.

Da sich die Äquivalenz der Items über die Wiederholbarkeit der Klassifikation bestimmt und über den konkreten Schüler der Teilpopulation objektiv vermittelt ist, läßt sich für den Vergleich und die Kombination der Items kein empirisch begründetes Meßinstrument zur Metrisierung angeben. Solange nur die erwähnte Teilpopulation zur versuchten Metrisierung der Items herangezogen wird, kann aufgrund der Gleichsetzung aller positiven Antworten darauf (dies erfolgt durch die Zuordnung der Ziffer 1 zu allen positiven Antworten) nur eine scheinbare Realisierung des Additivitätsprinzipes gegeben sein: Alle Items müssen als äquivalente, diagnostische Anregungsbedingungen betrachtet und können in Ermangelung eines populationsunabhängigen Meßinstrumentes nur zu einer Menge gleicher, abzählbarer Elemente kombiniert werden. Mit der Abzählbarkeit von Elementen verbindet sich keine Aussage über die empirischen Relationen zwischen ihnen.

Folglich stellt der Test als über die konkrete Schülerpopulation vermittelte Gesamtheit aller Items kein neues, in seiner empirischen Konstitution rekonstruierbares Objekt mit einer metrischen Struktur dar, sondern lediglich beziehungslose Aufeinanderfolge von Items. Die Herstellung der Beziehung zwischen den Items erfolgt empirisch durch das untersuchte Objekt "Schüler", in der Rekonstruktion aber nur durch den Forscher bzw. Diagnostiker.

In der Testtheorie wird nur die Rekonstruktion durch den Forscher problematisiert; sie vollzieht sich nach einer formalen und nicht interpretierbaren Veränderung des Prinzips der Additivität:

$$f(i_1 + i_2 + , \dots, + i_n) = f(i_1) + f(i_2) + , \dots, + f(i_n)$$

Durch die inadäquate Verwendung des +-Zeichens auf der linken Seite der Gleichung wird nichts anderes ausgedrückt, als daß die Anzahl der Funktionswerte der Items gleich der Anzahl der Items ist. Obwohl sich diese Aussage aus der Gleichung nicht ableiten läßt, muß sie aus der ungelösten Metrisierung und der praktizierten Anwendung von Tests erschlossen werden. Die linke Seite der Gleichung kann man nicht eindeutig interpretieren, weil das +-Zeichen in der gesamten Gleichung mehrdeutig ist. "Dieses Symbol bedeutet auf der rechten Seite der Gleichung etwas vollkommen anderes als auf der Linken Seite. Rechts ist es einfach ein Zeichen für die arithmetische Operation der Summenbildung: die Funktionswerte ... sind ja **Z a h l e n**, und diese sind zu addieren. Auf der linken Seite hingegen soll es eine bestimmte Art der Kombination physischer Objekte zur Bildung eines neuen physischen Objektes bezeichnen" (STEGMÜLLER, 1970, Band II, Studienausgabe Teil A, S. 51).

STEGMÜLLER will bei der Kritik an der obigen formalen Beziehung nur auf die fehlerhafte Schreibweise aufmerksam machen. In der Diagnostik hat diese Formalisierung eine reale, wenn auch widersprüchliche Entsprechung: Einerseits konstituiert sich der Funktionswert auf der linken Seite der Gleichung ausschließlich

durch die Summierung der Funktionswerte auf der rechten Seite. Der Operator "+" wird dadurch funktionslos. Auf der anderen Seite kann man aber die Summierung nur dann legitimieren, wenn die Items in unmittelbarer Aufeinanderfolge dargeboten und bearbeitet werden. Dies impliziert wiederum die Befolgung der Darbietungs-, Bearbeitungs- und Auswertungsregeln der Testdurchführung und damit die Gegebenheit von Standards, unter denen sich den Items allgemeine Funktionswerte zuordnen lassen. Werden solche Standards nicht eingehalten, dann verändern sich die bekannten Funktionswerte der Items wie Lösungswahrscheinlichkeit, Trennschärfe, Reliabilitätsindex, Validitätsindex usw. (vgl. MICHEL, 1964). Ihre arithmetische Kombination auf der rechten Seite der Gleichung würde auf der linken Seite zu unterschiedlichen Funktionswerten für den Gesamttest führen. Aus diesem Grund repräsentiert hier der Operator "+" die Regeln zur Durchführung und Bearbeitung des Tests und fungiert als Vorschrift, die Bearbeitung aller Items organisatorisch zu kombinieren.

Das +-Zeichen als Operator auf der linken Seite der Gleichung symbolisiert die Vorschrift, die Entwicklung und Anwendung eines Beurteilungsverfahrens zu standardisieren, ohne die Auswirkung der Standardisierung auf die individuumsspezifische Auseinandersetzung mit dem Test explizit in Rechnung zu stellen. Der Vorgang der Standardisierung erlaubt keinerlei Zuordnung von Funktionswerten zu verschiedenen Items; er stellt kein Meßinstrument für ihren Vergleich und ihre Kombination dar. Da die physikalischen Eigenschaften der Items oder die zwischen ihnen bestehenden, physikalisch beschreibbaren Relationen für ihre Beantwortung irrelevant sind bzw. als irrelevant betrachtet werden, verbleibt nur die Möglichkeit, das Resultat ihrer Beantwortung durch die Schüler für die Bestimmung der Itemcharakteristiken heranzuziehen. Die Kombination von Items zu einem Gesamttest, d.h. ihre Auswahl und Anordnung, läßt sich damit nur durch die Legitimation solcher Itemcharakteristiken und ihrer rechnerischen Synthese darstellen. Der Funktionswert eines

Tests als angeblich empirisch bedeutungshaltige Kombination verschiedener Items erschöpft sich daher in der arithmetischen Kombination der Funktionswerte der Items und wird nicht unabhängig von innen gewonnen. Er kann die Funktionswertbestimmung der Einzelitems im empiristischen Sinne empirisch nicht bestätigen.

Zusammenfassend gesehen, scheitert die Interpretation von "Meßwerten" in der Schülerbeurteilung an der empirischen Nichtdeutbarkeit des Kombinationsoperators. Der Operator "o" läßt sich in der angeführten Regel für die Einführung von metrischen Begriffen, dem Additivitätsprinzip, empiristisch nicht rekonstruieren, da man keine Kombinationsmethode für die Einzelitems angeben kann, die eine empirische Relation zwischen den Items hypothetisch oder gesetzesartig verallgemeinert. Der eigentlich empirisch bedeutungshaltige Operator für die Kombination der Items ist das untersuchte Individuum, das sich mit den verschiedenen Items auseinandersetzt, Antworten produziert und dabei das Beurteilungsverfahren in seinem Zusammenhang konstituiert. Es würde damit als "Meßinstrument" für die Metrisierung der Einzelitems fungieren; tatsächlich wird aber der Test, für den noch keine metrisierte Dimension existiert, als "Meßinstrument" für die Bestimmung von quantitativen Größenwerten, den Ausprägungsgraden dispositioneller Eigenschaften der untersuchten Individuen, verwendet.

Bei der Diskussion des SCHMIDT'schen Rahmenmodells für die Rekonstruktion dispositioneller Eigenschaften wurde die praktizierte Diagnostik in der Schule im allgemeinen dadurch charakterisiert, daß die Diagnostiker- und Forschergemeinschaft die Reflexion des empirischen Geschehens im Interaktionssystem des untersuchten Individuums, S_I , vernachlässigt. In Bezug auf den Versuch der Metrisierung dispositioneller Eigenschaften äußert sich der Rückzug auf das Interaktionssystem des Forschers in der Verkennung und Vernachlässigung der objektiven Funktion des untersuchten Individuums als empirischer Operator für die Kombination der verschiedenen Items.

2.4.5.4 Die Entsubjektivierung des Deutungsproblems

Nach dem vorliegenden Ergebnis der Diskussion muß man feststellen, daß in der empiristisch orientierten Diagnostik in der Schule **d i e empirisch objektive Grundlage**, der Aufgaben beantwortende Schüler, in die Entwicklung von diagnostischen Modellen nicht miteinbezogen wird. Der Ausprägungsgrad dispositioneller Eigenschaften bemißt sich nach der Anzahl der produzierten, positiven Lösungsalternativen. Da es für die Bewertung dispositioneller Eigenschaften nur auf die Anzahl der positiven Lösungsalternativen pro untersuchten Individuum ankommt, die spezifische Qualität der Alternativen sowie der sie provozierenden Items also irrelevant ist, müssen die verschiedenen Items als untereinander austauschbar bzw. äquivalent betrachtet werden. In Ermangelung eines empirisch bedeutungshaltigen Meßinstrumentes zum Vergleich der Items kann daher die Behauptung, daß sie äquivalent seien, nur indirekt legitimiert werden: durch den Nachweis, daß verschiedene Individuen zu verschiedenen Zeitpunkten die gleiche **A n z a h l** von positiven Lösungsalternativen produzieren.

Erfüllt sich die Erwartung unter Verwendung des gleichen Tests, so spricht man von seiner Reliabilität, unter Verwendung eines davon verschiedenen Tests oder Beurteilungsverfahrens von seiner Validität. Beide Begriffe implizieren Stabilitätskennzeichen von Tests und drücken somit Invarianzmomente des Beurteilungsprozesses aus. Die Invarianz sollte sich bei der Anwendung der Verfahren in der Produktion gleicher Anzahlen von positiven Lösungsalternativen pro Schüler äußern. Zeigt sich dieser Effekt tatsächlich, dann wird die Äquivalenz indirekt durch die falsche Interpretation des Prinzips der Additivität über die Summierung der Antworten **d e f i n i t o r i s c h** festgelegt. Jedes Testitem kann den gleichen Beitrag zur Erhöhung des Summenwertes leisten. Auch bei unterschiedlicher Gewichtung der verschiedenen Items ändert sich an der definitiven Festlegung nichts, weil für die Gewichtung selbst kein

empirisch bedeutungshaltiges und populationsunabhängiges Meßinstrument für ihren Vergleich zur Verfügung steht.

Die definitorische Festlegung der Itemsäquivalenz über das invariante Auftreten gleicher Anzahlen von positiven Lösungsalternativen gilt allerdings nur in einem relativen Sinn, der im folgenden erläutert werden soll. Denn der konkrete Versuch, die Annahme zu bestätigen, daß auf die wiederholte Darbietung desselben oder eines strukturell gleichen Items die gleiche Reaktion erfolgt, ist sinnlos weil tautologisch, solange keine reaktionsunabhängige Bestimmung des Items und eine itemunabhängige Bestimmung gleicher Reaktionen vorliegt. Man kennt dieses Problem als Schwierigkeit in der Konstruktion und Anwendung von Tests und es wurde durch die wissenschaftliche Untersuchung des Lernzuwachses in gleichen Problemsituationen hinreichende erforscht (vgl. insbesondere zum Effekt der Übung FOPPA, 1965). Diese Untersuchungen zeigen, daß die Versuchspersonen ihre aktive Auseinandersetzung mit gleichen oder ähnlichen Problemsituationen in der Zeit ändern.

Aufgrund der Veränderung der Verhaltensprodukte in der Zeit wäre somit die definitorische Festlegung der Itemsäquivalenz nicht gerechtfertigt. Außerdem verändern sich in der Zeit die situativen Bedingungen der Itemdarbietung und -bearbeitung, sodaß auch von daher das invariante Auftreten gleicher Anzahlen von positiven Lösungsalternativen für jedes untersuchte Individuum nicht die Annahme der Itemsäquivalenz legitimieren könnte. Realiter ist zudem das Auftreten gleicher Anzahlen von positiven Lösungsalternativen pro untersuchtem Individuum nicht feststellbar. In der Testtheorie versucht man diesem Problem durch die Berücksichtigung und Kontrolle des "Meßfehlers" gerecht zu werden. Die Handhabung des Meßfehlerkonzeptes bei der Konstruktion von Tests schafft - insbesondere in der klassischen Testtheorie - die Voraussetzung für die definitorische Festlegung der Itemsäquivalenz in dem zu erläuternden, relativen Sinn.

2.4.5.4.1 Verfahrenstechnische Festlegung der Itemäquivalenz

In Analogie zum Vorgehen in den Naturwissenschaften stellt der Meßfehler jene variable Größe dar, die bei wiederholter Anwendung eines Tests unter kontrollierten gleichen Bedingungen als störende Schwankung des Zeigerausschlages, hier als Varianz in der Anzahl der von einem bestimmten untersuchten Individuum produzierten, positiven Lösungsalternativen, in Erscheinung tritt. Der Meßfehler ist das Komplement zur unterstellten Invarianz des Zeigerausschlages, die als wahrer Wert und als repräsentativ für den Ausprägungsgrad einer dispositionellen Eigenschaft des untersuchten Individuums aufgefaßt wird. Da der Test lediglich die Schwankungen indiziert und den hypothetischen, wahren Wert nicht unmittelbar zu erkennen gibt, besteht die größte Schwierigkeit in der Testtheorie darin, den hypothetischen Wert aus den indizierten Werten zu rekonstruieren und die Annahmen für die Rekonstruktion zu begründen.

"Der wahre Meßwert (wird; d. Verf.) als der Erwartungswert einer Variablen X_{vi} eingeführt, welche sich semantisch als die unendlich groß gedachte Menge möglicher Testwiederholungen unter gleichen Bedingungen interpretieren (läßt; d. Verf.). Klarerweise hängt ihre Verteilung wesentlich von ... den in der Testvorgabe konstant gehaltenen Bedingungen (ab; d. Verf.). Im allgemeinen wird eine Testleistung der Vpn unter Ermüdung schlechter sein als im ausgeruhten Zustand, in ruhiger Umgebung besser als bei Lärm, und so weiter. Wird der experimentelle Faktor "Ermüdung" konstant gehalten, ergibt sich eine geringere Meßfehlervarianz als dann, wenn die Ermüdung der Vp nicht kontrolliert wird. Verschiedene Versuchsbedingungen bewirken daher verschiedene "wahre" Werte und unterschiedliche Fehlervarianzen" (FISCHER, 1974, S. 32).

FISCHER weist darauf hin, daß dieser Umstand zu einer berechtigten Kritik an der klassischen Testtheorie geführt hat und daß sich die Grundlage der Kritik auf die Problematik einer wirklich äquivalenten Messung zurückführen lassen. Ob andere testtheoretische Modelle diese Problematik bewältigen können, kann hier nicht diskutiert werden. Überdies muß für die Betrachtung der praktizierten Diagnostik in der Schule die Orientierung an der klassischen Testtheorie im Vordergrund der Diskussion

stehen, weil sie der Form der selektionsbezogenen Schülerbeurteilung am besten entspricht.

Nach dem bisher Gesagten hängt die Rekonstruktion des hypothetischen Wertes aus den indizierten Werten von der Konstanthaltung und Kontrolle verschiedener Versuchsbedingungen ab. Ohne ihre experimentell Kontrolle erhöht sich die Ungenauigkeit der Beurteilung. Berücksichtigt man nun die Tatsache, daß der Zeigerausschlag des Tests, soweit er formal rekonstruierbar ist, das Resultat der rechnerischen Synthese der durch Zahlen symbolisierten Antworten auf die verschiedenen Testitems darstellt und die rechnerische Synthese mit hinreichender Zuverlässigkeit wiederholt werden kann, so ist die als Meßfehler interpretierbare Schwankung des Zeigers auf die variable Bearbeitung der Testitems durch das untersuchte Individuum in der Zeit zurückzuführen. Die Ursachen dafür können auf aktuelle Befindlichkeiten des untersuchten Individuums - der erwähnte Ermüdungszustand - oder situative Bedingungen wie störender Lärm zurückführbar sein. Solche Störfaktoren lassen sich grundsätzlich, d.h. soweit sie bekannt sind, durch die sukzessive Standardisierung der diagnostischen Situation konstant halten und kontrollieren. Das Lösungsverhalten wird auf diese Weise zunehmend durch Durchführungs- und Darbietungsregeln für den Test funktionalisiert. Die so vorgenommene Standardisierung der diagnostischen Situation ist jedoch nur insofern von Bedeutung, als dadurch für alle untersuchten Individuen die gleichen Reaktionsbedingungen zu schaffen versucht werden.

Demgegenüber interessiert aber hier die Frage, ob für jedes einzelne, untersuchte Individuum die den Test konstituierenden einzelnen diagnostischen Anreizbedingungen, die verschiedenen Items, als Meßwiederholungen aufzufassen sind. Schließlich ergibt sich die Äquivalenzforderung an die verschiedenen Items aus dem Vorgehen, die ihnen entsprechenden und vom untersuchten Individuum produzierten, positiven Lösungsalternativen als gleichwertige zu einem Summenwert zusammenzufassen.

In Bezug auf die verschiedenen Ursachen von Meßfehlern müßten die Items hinsichtlich dreier Aspekte für das *e i n z e l - n e* , untersuchte Individuum äquivalent sein:

1. Die verhaltensrelevanten, physikalisch beschreibbaren Aspekte der Items dürften sich nicht verändern.
2. Die in den Items enthaltene und abzubauen Information müßte in allen aufeinanderfolgenden Items die gleiche sein.
3. Die subjektive Befindlichkeit des untersuchten Individuums müßte während der gesamten Testbearbeitung unverändert bleiben.

In der Terminologie des Interaktionsmodelles von SCHMIDT (1974) bedeutet dies, daß die Überprüfung der Itemäquivalenz nur dann möglich ist, wenn dem Forscher bzw. Diagnostiker differenzierte Kenntnisse über die Rekonstruktion der diagnostischen Situation durch das untersuchte Individuum vorliegen. Zudem müßten sich die Kenntnisse in Form von Naturgesetzen darstellen lassen, da sie eine invariante Beziehung zwischen der Reizlage und den Verhaltensprodukten verallgemeinern sollen. Daß die Beschränkung der naiv-empiristischen Forschung auf die Vorgänge im Interaktionssystem des Forschers die Gewinnung solcher Kenntnisse nicht zuläßt, wurde bereits ausführlich diskutiert. Daß "die Äquivalenzannahmen der klassischen Testtheorie ... psychologisch unplausibel und praktisch unrealisierbar" (FISCHER, 1974, S. 26) sind, ist in der neueren testtheoretischen Literatur ebenfalls hinreichend belegt. Wie wird aber in der praktizierten Diagnostik trotz Nichterfülltheit der Bedingungen die Annahme der Itemäquivalenz legitimiert ?

Ausgehend von der Funktion eines Tests, zwischen den Individuen einer Population optimal zu unterscheiden, kann man sich nicht mehr dafür interessieren, ob die Testwiederholung für jedes untersuchte Individuum zum gleichen Ergebnis führt. Die Forderung nach der Reproduktion des gleichen Zeigerauslasses mittels Test läßt sich nicht mehr aufrechterhalten, weil die Bearbeitung jedes einzelnen Items des Tests durch das untersuchte Individuum unterschiedlich ist und jeweils unterschiedliche,

wahre Werte angenommen werden können. Auch die wiederholte Vorlage einunddesselben Items führt zu Veränderungen zu Reaktionen in der Zeit, wie z.B. durch die Untersuchungen zur "semantischen Sättigung" nachgewiesen werden konnte (vgl. POLLIO, 1968).

Stattdessen versucht man Tests in der Weise zu konstruieren, daß die Verhältnisse zwischen den untersuchten Individuen einer Population invariant bleiben. Die Verhältnisse zwischen den untersuchten Individuen treten bei Anwendung eines Tests in der Varianz der individuumspezifischen Summenwerte in Erscheinung.

Läßt sich diese Varianzstruktur bei der wiederholten Durchführung des gleichen Tests oder bei der Darbietung eines ähnlichen Items reproduzieren, so wird in Bezug auf die Anwendungspopulation von der Äquivalenz der Tests oder allgemeiner, von ihrer Zuverlässigkeit gesprochen. Entscheidend ist dabei, daß sich die numerischen Ergebnisse im ersten Test in die des zweiten Tests durch eine mathematische Funktion transformieren lassen und umgekehrt. Aus diesem Grund können die absoluten Summenwerte für jedes Individuum in beiden Tests jeweils unterschiedlich sein, das Verhältnis der individuumspezifischen Summenwerte zueinander muß aber beim Übergang vom ersten zum zweiten Test konstant bleiben. Da sich die Summenwerte nur in ihrer formalen rechnerischen Synthese rekonstruieren lassen und die für sie konstitutiven, positiven Lösungsalternativen untereinander austauschbar sind, kann die objektive Bedingung für die festgestellte Invarianz des Verhältnisses zwischen den Summenwerten nur in den individuumunabhängigen Gemeinsamkeiten des Entstehungsprozesses zu suchen sein, der sich von der Itemdarbietung über ihre Bearbeitung bis zur Klassifikation der Verhaltensprodukte erstreckt. Das heißt: Die festgestellte Invarianz kann nur in dem Ausmaß auftreten, wie der Beurteilungsprozeß im Interaktionssystem S_T reglementiert ist.

Die Forderung nach Äquivalenz der Items bezieht sich somit - und das ist mit Äquivalenz im relativen Sinne gemeint - auf deren gemeinsame Fähigkeit zur Unterscheidung der Summenwerte beizutragen. Die Funktion der Zusammenfassung verschiedener Items zu einem Gesamttest kann daher nicht darin bestehen, die Invarianz von individuumspezifischen Verhaltenstendenzen nachzuweisen, sondern nur die Tatsache wiederzugeben, daß es eine diagnostische Methode gibt, mit der die Stabilität individueller Unterschiede in der Zeit zu beschreiben ist. Durch diese Methode, die im Bereich praktizierter Diagnostik in der Schule durch die klassische Testtheorie systematisiert wird, soll die Betragsleistung der Items eines Tests zur Diskriminierung zwischen den Schülern charakterisiert werden.

Von jedem Item wird verlangt, daß es die einzelnen Personen der untersuchten Population durch seine Beantwortung in das abstrakte Verhältnis von Löser und Nichtlöser stellt. Zeigt sich bei der wiederholten Darbietung der gleichen Items oder ähnlicher Items der gleiche Effekt, so gilt die diagnostische Prozedur als zuverlässig. Konstitutiv für die Feststellung der Invarianz des Verhältnisses zwischen den untersuchten Individuen sind drei Bedingungen:

1. Die Wiederholung der gleichen Klassifikation der Verhaltensprodukte, die als Antworten auf die Darbietungen der gleichen Items oder auf die Darbietung ähnlicher Items geliefert werden.
2. Die optimale Standardisierung des Darbietungsprozesses zur Schaffung gleicher Anregungsbedingungen. Bei der Darbietung der ähnlichen Items ist vorauszusetzen, daß ihre Unterschiede keine Bedeutung für die Produktion der Lösungsalternativen haben, d.h., es muß eine lösungsrelevante Definition der Ähnlichkeit vorliegen.
3. Die Bearbeitung der wiederholten oder ähnlichen diagnostischen Anregungsbedingungen durch die gleiche Population. Ansonsten ist der Koeffizient für die Lösungswahrscheinlichkeit nicht zu interpretieren.

In allen drei Fällen lassen sich die Bedingungen durch reine Festlegung erfüllen; ihre Gegebenheit kann somit ohne Bezug auf eine empirische Überprüfung angenommen werden.

Für die Erfüllung der Bedingungen muß die Gemeinschaft der Diagnostiker und Forscher die Grenzen des Interaktionssystems S_T nicht übertreten; sie kann ihre Reflexionsbemühungen auf die Reglementierung der Klassifikation der Verhaltensprodukte, die optimale Standardisierung des Darbietungsprozesses und die Auswahl der Untersuchungspopulation beschränken. Begnügt man sich im Rahmen der Schülerbeurteilung mit diesem Vorgehen, so kann man damit nicht mehr als das feststellbare Phänomen beschreiben, daß die dargebotenen Items hinsichtlich der erfolgten Klassifikation der Verhaltensprodukte, der Spezifität der Standardisierung des Darbietungsprozesses und der Spezifität der untersuchten Population äquivalent sind. Ihre Gleichheit bzw. Invarianz ist damit legitimiert aus der Invarianz des Verhältnisses von Lösern und Nichtlösern, für dessen Feststellbarkeit die Gleichheit der Klassifikation der Verhaltensprodukte, der Standardisierung des Darbietungsprozesses und der untersuchten Population anzunehmen ist. Bei Änderung irgendeines dieser Bedingungskomplexe ist die Äquivalenzannahme für die Items nicht mehr aufrechtzuerhalten.

Aus diesem Grund ist sie nur in dem Ausmaß gültig, wie die festgelegten Bedingungen eindeutig bestimmt sind und beim Vergleich der Items explizit benannt werden. Und nur dann lassen sich Items als Wiederholungen eines diagnostischen Anregungssituation begreifen, mit der ein invariantes Verhältnis zwischen Lösern und Nichtlösern ausgewiesen werden soll. Wenn man von den realiter gegebenen und auch intendierten Unterschieden in der Lösungswahrscheinlichkeit der einzelnen Items in einem Test vorerst absieht und den Test als eine komplexe, diagnostische Anregungsbedingung begreift, die sich aus einer Menge von Äquivalenten Items in diesem relativen Sinn zusammensetzt, dann ist das Gemeinsame der Items nur Ausdruck der auf definitionischem Weg festgelegten, konstitutiven Bedingungen.

Mit dem Versuch der Ermittlung der relativen Itemäquivalenz und dem praktizierten Vorgehen, die Items zu einem Test zu Kombi-

nieren - von den unterschiedlichen Lösungswahrscheinlichkeiten im konkreten "Test"-fall soll hier wiederum abgesehen werden -, vollzieht sich die Überprüfung von zwei Voraussetzungen für die Beurteilung der untersuchten Individuen, die sich gegenseitig bedingen und im Forschungsprozeß nicht voneinander zu trennen sind:

1. die Invarianz individuumspezifischer Lösungsprozesse und die Konstanz interindividueller Unterschiede;
2. die relative Itemäquivalenz in dem Sinn, daß die Beantwortung der Items in allen Fällen zur gleichen Klassifikation der Population führt.

Durch die Überprüfung beider Voraussetzungen mit einunddemselben Vorgehen wird der Untersuchungsprozeß tautologisch; es soll damit gleichzeitig die Adäquanz der vorgenommenen Personenkennzeichnung und die der methodischen Hilfsmittel zur Personenkennzeichnung nachgewiesen werden. In der Kontamination beider Voraussetzungen bzw. ihrer Überprüfung kommt somit wieder jener naiv-empiristische Ansatz zum Ausdruck, in dem zwischen Gegenstandsbestimmung und empiristischer Bestätigung nicht unterschieden wird (DICK, 1974). Für den Nachweis der relativen Itemäquivalenz muß deshalb die Konstanz individueller Unterschiede als quasi naturgegeben vorausgesetzt werden. Letzteres zeigt sich aber immer nur an der relativen Äquivalenz der Items, deren populationsunabhängige Gemeinsamkeiten definitionell festgelegt werden. Sie konkretisieren sich in dem Versuch der optimalen Standardisierung der gesamten diagnostischen Situation, für die die Diskriminierung der Population in Löser und Nichtlöser, unabhängig von der empirischen Untersuchung des diagnostischen Prozesses sowie des Lösungsprozesses, in der praktizierten Diagnostik in der Schule vorgegeben ist. Somit ist das Diskriminationsprinzip für den Vergleich von Items und für die Festlegung der relativen Itemäquivalenz konstitutiv und entscheidet über die Brauchbarkeit bzw. Unbrauchbarkeit eines Items in einem Test.

Die Realisierung des Diskriminationsprinzips in einem konkreten Test und in Bezug auf die Itemäquivalenz in Form des wiederholten Auftretens von gleichen Lösungswahrscheinlichkeiten drückt also nur aus, ob der diagnostische Prozeß hinreichend reglementiert ist, um durch alle Items die Population in gleicher Weise zu klassifizieren. Die empirische Bedeutung des Funktionswertes der verschiedenen Items, der Lösungswahrscheinlichkeit, ist unbekannt. Die Items sind nur hinsichtlich des Verhältnisses zwischen Lösern und Nichtlösern äquivalent, nicht aber hinsichtlich des Auseinandersetzungsprozesses in den untersuchten Individuen.

Realiter wirken sich die Darbietungen und Bearbeitungen der einzelnen Testitems nicht als Wiederholungen gleicher diagnostischer Anregungsbedingungen aus. Die unterschiedliche Aufteilung der untersuchten Population in zwei Teilklassen durch die verschiedenen Items werden auf die sog. Meßfehler zurückgeführt, deren Ursache in der Nichterfülltheit der für die Invarianz des Verhältnisses notwendigen Bedingungen zu suchen ist. Die Ungenauigkeit in der Wiederholung der gleichen Klassifikation der Verhaltensprodukte wird als mangelnde Objektivität bezeichnet, die in der Wiederholung des Darbietungsprozesses als unzureichende Reliabilität des "Meßinstrumentes" und die in der Wiederholung der Population mit dem Begriff der "Fluktuation" des sog. wahren Wertes der dispositionellen Eigenschaft. Aus diesem Grund werden die einzelnen Items zu einem Gesamttest zusammengefaßt. Der Zweck der Zusammenfassung besteht in der Kompensation der verschiedenen Fehlerkomponenten. Die Annahme, daß die Zusammenfassung der Items zu einem Gesamttest zur Kompensation der Fehlerwerte führt, findet ihren Ausdruck in der rechnerischen Synthese der Funktionswerte der Einzelitems zu einem Funktionswert des Gesamttests.

Dieser Funktionswert, die Reliabilität des Tests, beansprucht ein Maß für die Wiederholbarkeit der Itemfolge als Gesamtes bzw. der Itemkombination zu sein. Für den vorliegenden Fall

eines Tests, dessen Items zu einer annähernd wiederholbaren Aufteilung der untersuchten Population in Löser und Nichtlöser beitragen sollen, berechnet sich der Funktionswert der Reliabilität für den Gesamttest nach der Formel von KUDER-RICHARDSON (1937):

$$\text{Rel.} = \frac{n}{n-1} \left(1 - \frac{\sum p_i(1-p_i)}{s_x^2} \right)$$

Rel. = Reliabilität, p_i = Lösungswahrscheinlichkeit der Items i , s_x^2 = Varianz der Summenwerte in der untersuchten Population, n = Anzahl der Items.

In der Formel für die Reliabilitätsschätzung nach KUDER-RICHARDSON gelten die einzelnen Items explizit als Wiederholungen des gleichen Meßvorganges. Auf die Reliabilitätsschätzung für Tests mit Items von unterschiedlicher Lösungswahrscheinlichkeit, d.h. für Tests, deren Einzelitems nicht als Wiederholungen aufzufassen sind, wird später eingegangen.

Die Formel ist ein Ausdruck der durchschnittlichen Korrelation zwischen den einzelnen Aufgaben eines Tests, von denen angenommen wird, daß sie inhaltlich gleich sind und gleiche Interkorrelationen haben. Unter Gültigkeit dieser Voraussetzungen und der Voraussetzung der gleichen Lösungswahrscheinlichkeit der Aufgaben strebt die Reliabilität gegen einen Maximalwert. Es läßt sich ableiten, daß die Formel die Verrechnung der Reliabilitäten der Einzelitems im Test darstellt (vgl. MAGNUSSON, 1969, S. 124).

In dem Maße wie die Einzelitems den beiden Forderungen der gleichen Lösungswahrscheinlichkeit und der gleich großen Interkorrelationen nicht genügen, nimmt die Reliabilität des Gesamttests ab. Das bedeutet, daß die einzelnen Items entweder die untersuchte Population nicht in Klassen aufteilen, deren Umfang für alle Items gleich groß ist; es kann aber auch bedeuten, daß trotz gleicher Lösungswahrscheinlichkeit für alle Items die Individuen in den Klassen jeweils nicht identisch

sind. Die beiden unerwünschten Effekte lassen sich zwar unabhängig voneinander überprüfen - ersterer durch den Vergleich der Lösungswahrscheinlichkeiten, der zweite durch die Berechnung der Interkorrelationen zwischen den Testaufgaben -, ihre Interpretation als der verursachende Störfaktor ist aber nicht möglich, da sie voneinander abhängen. Dies liegt an der Determiniertheit der Interkorrelationen durch die Lösungswahrscheinlichkeit der Items. Die Interkorrelationen der Items sind in der obigen Formel für die Bestimmung der Reliabilität des Gesamttests im Ausdruck für die Varianz enthalten und lassen sich als arithmetischer Zusammenhang der Lösungswahrscheinlichkeiten rekonstruieren.

Damit ist aber die Reliabilität des Gesamttests ausschließlich das Produkt der Verrechnung der Lösungswahrscheinlichkeiten der verschiedenen Items. Es wird an keiner Stelle der Verrechnung auf Itemkennzeichen Bezug genommen, die sich aus einem empirischen Vergleichsmaßstab für die Items ergeben. Eine empirische Operation für die Zuordnung des Funktionswertes "Reliabilität" zum Gesamttest läßt sich in Entsprechung zur Realisierung des Additivitätsprinzipes für die Einführung einer metrischen Größe nicht angeben. Der Funktionswert der Reliabilität ist ein auf arithmetischem Weg gewonnener Durchschnittswert, der die verschiedenen Beiträge der Einzelitems zur Diskrimination der untersuchten Individuen auf der Grundlage der Testsummenwerte verallgemeinert. Das, was allen Items in Bezug auf die Diskriminationsfähigkeit gemeinsam ist, bestimmt sich aus dem Verhältnis der Summe der Unterscheidungsfähigkeiten der Einzelitems bzw. der durch sie erzeugten Varianz ($\sum p_i(1-p_i)$) zur Unterscheidungsfähigkeit bzw. Varianz, die durch den Gesamttest bedingt ist (s_x^2).

Obwohl über die empirische Bedeutung der dahinterstehenden Unterscheidungsdimension noch nichts ausgesagt ist, fungiert die Höhe der Reliabilität des Gesamttests, also der Ausdruck dafür, in welchem Ausmaß die verschiedenen Items als Wiederholung einer diagnostischen Anreigungsbedingung aufgefaßt werden sollen,

als Kriterium der Nützlichkeit eines Beurteilungsverfahrens. Die inhaltliche Bedeutung des Wiederholten kann erst nach diesem Nützlichkeitsbeweis zu bestimmen versucht werden. Deswegen ist die Größe, die man in der Testtheorie als Meßfehler bezeichnet, lediglich ein quantifizierter Ausdruck für die Nichtwiederholbarkeit der Folge von dargebotenen Items als diagnostische Anregungsbedingungen und spiegelt somit die ungenügende Standardisierung des objektivierten diagnostischen Prozesses wider.

Die definitorische Festlegung der Standards bezog sich ja auf die Klassifikation der Verhaltensprodukte, die optimale Vereinheitlichung des Darbietungsprozesses und die Identität der untersuchten Population. Sofern man in der Klassifikation der Verhaltensprodukte ein intersubjektiv konsistentes Vorgehen durch die Beurteiler voraussetzen kann, indem man nur solche Items in einem Test zuläßt, für die die positive Lösungsalternative eindeutig bestimmt ist, und sofern man die Identitätsbedingung erfüllt, kann die Fehlerhaftigkeit in der Wiederholung der diagnostischen Anregungsbedingung nur noch auf sie selbst zurückzuführen sein. Das heißt: Die durch die verschiedenen Items provozierten Auseinandersetzungen des untersuchten Individuums sind in ihrer Komplexität noch zu verschieden, um jedesmal in dieselbe Klasse der Löser bzw. Nichtlöser sortiert zu werden. Es gilt daher die verschiedenen diagnostischen Anregungsbedingungen untereinander noch weiter zu vereinheitlichen. Die dazu erforderlich Instrumentalisierung des Testverhaltens erstreckt sich von der Kontrolle und Reglementierung situativer Bedingungen der Testdurchführung über die Herstellung formaler Ähnlichkeiten zwischen den verschiedenen Items bis zur Kontrolle und versuchten Manipulation innerpsychischer Prozesse, deren Einfluß auf den Verlauf des Testverhaltens auf psychometrischem Weg ermittelt wurde (vgl. MICHEL, 1964). Kriterium für die Zulässigkeit solcher Instrumentalisierungen ist die Erhöhung der Reliabilität, d.h. die Herstellung eines invarianten Verhältnisses zwischen Lösern und Nichtlösern in der

untersuchten Population über alle Items.

Ob die verschiedenen Items als Wiederholungen diagnostischer Anreizbedingungen aufgefaßt werden können, bemißt sich im Sinne ihrer relativen Äquivalenz nach ihrer Gleichheit in der Unterscheidung der Individuen der untersuchten Population. Da es die untersuchten Individuen sind, die die verschiedenen Items durch sukzessive Bearbeitung als unterstellte Wiederholung der gleichen Lösungsprozesse empirisch miteinander kombinieren, werden sie indirekt zum empirischen Kriterium für die konstatierte, relative Invarianz des Verhältnisses von Lösern und Nichtlösern. Sie konstituieren somit die feststellbare Reliabilität des Tests als zusammenhängende Gesamtheit aller Einzelitems.

Die Verwendung des Tests als reliables Instrument zur Unterscheidung von Individuen bei einer anderen Population ist infolgedessen nicht zulässig. Desgleichen kann ebensowenig von der Reliabilität eines Tests gesprochen werden, da ihre Höhe von der Spezifität der Population abhängig ist. Sie erfährt ihre Existenzberechtigung nur im Rahmen einer bestimmten Standardisierung des diagnostischen Prozesses und nur innerhalb einer bestimmten Untersuchungspopulation. M.a.W.: Ob ein Test unter dem Gesichtspunkt der Reliabilität auf andere empirische Bedingungen der Konstitution von Lösungsergebnissen, nämlich auf die situativen Bedingungen, unter denen der diagnostische Prozeß abläuft, und auf eine andere Untersuchungspopulation bzw. die gleiche Untersuchungspopulation zu einem anderen Zeitpunkt übertragbar ist, muß im konkreten Fall jeweils ausgewiesen werden. Man kann es nur nach der vollzogenen Übertragung entscheiden.

Damit kann aber der Funktionswert der Reliabilität eines Tests nicht den Anspruch erheben, ein empirisch begründetes, metriertes Kennzeichen des Tests zu sein. Er ist nur als Resultat der arithmetischen Kombination der Lösungswahrscheinlichkeiten der Einzelitems darstellbar und ist deshalb ein ausschließlich auf analytischem Weg gewonnenes Kennzeichen der als Darbietungsfolge zusammengefaßten Items.

2.4.5.4.2 Die Formalisierung der Deutung von Schülerbeurteilungen

Bei der Untersuchung des Vorgehens, wie in der Objektivierung der Schülerbeurteilung diagnostische Anregungsbedingungen als Bemessungsvorschriften für die Ermittlung des Ausprägungsgrades dispositioneller Eigenschaften festzulegen versucht werden, wurde bis jetzt nur eine spezifische Form der Reliabilitätsbestimmung von Beurteilungsverfahren betrachtet, die Einschätzung der internen Konsistenz eines Tests nach KUDER-RICHARDSON. Der Anwendung dieses Bestimmungsverfahrens liegt die Frage zugrunde, inwieweit die verschiedenen Items in einem Test als Wiederholung der gleichen diagnostischen Anregungsbedingung betrachtet werden können. Die Frage ergibt sich aus den Problemen, die mit der versuchten Metrisierung von Dispositionsbegriffen in der empiristisch orientierten Diagnostik verbunden sind. Erst wenn es gelänge, die verschiedenen Items in einem Test als Wiederholungen der gleichen diagnostischen Anregungsbedingung empirisch zu begründen, wäre die Voraussetzung dafür geschaffen, die Kombination solcher Items zu einer Aufeinanderfolge in einem Gesamttest unter dem Gesichtspunkt zu untersuchen, ob unterschiedlich häufige Beantwortungen unterschiedliche Ausprägungsgrade einer dispositionellen Eigenschaft repräsentieren. Die Analyse zeigte, daß der Nachweis der Itemäquivalenz nur in einem relativen Sinne unternommen werden kann: Items können nur unter dem Aspekt als äquivalent erscheinen, daß sie die untersuchte Population in gleicher Weise in die Klasse der Löser und die der Nichtlöser aufteilen. Da diese Art der Unterscheidungsfähigkeit eine notwendige Vorbedingung für die Einschätzung ihrer Brauchbarkeit und für ihre Verwendung ist, wird die Itemäquivalenz auf definitorischem Weg festgelegt und nicht empirisch begründet.

Die Priorität der Wirksamkeit des Diskriminationsprinzips für die Festlegung der Gleichwertigkeit von Testaufgaben liegt in der Selektionsfunktionalität von Urteilen begründet.

Sie erweist sich bei der Bestimmung der Itemäquivalenz als Hindernis für die empiristische Einführung von Eigenschaftsbegriffen als metrische in eine intersubjektiv konsistent gebrauchte diagnostische Sprache. Die notwendige Einführung von metrisierten Begriffen in die Wissenschaftssprache wird daher in der Diagnostik in der Schule durch die Auszählung der Menge von richtig beantworteten Testaufgaben umgangen. Darin manifestiert sich eine falsche Anwendung und Interpretation des Additivitätsprinzips.

Darüberhinaus legitimiert das Diskriminationsprinzip die Festlegung von unterschiedlich wertigen Testitems bei der Bestimmung der Reliabilität und Validität von zusammenhängend dargebotenen Itemfolgen, d.h. von Test in ihrer Itemgesamtheit. Die hier zu Anwendung kommenden Verfahren zur Funktionswertbestimmung von Tests, die weitgehend auf der Formalisierung der klassischen Testtheorie durch GULLIKSEN (1950) beruhen, sind in der Praxis nur dann einsetzbar, wenn die verschiedenen Testitems auf unterschiedlichen Anforderungsniveaus die untersuchte Population jeweils in die zwei Klassen der Löser und Nichtlöser einteilen, d.h. wenn die verschiedenen Testitems unterschiedliche Lösungswahrscheinlichkeiten haben, die ausschließlich über das formale Verhältnis zwischen Lösern und Nichtlösern bestimmt werden. Der Effekt dieses Vorgehens äußert sich in der Vergrößerung der Gesamtvarianz des Tests und somit in der häufigeren und differenzierteren Diskrimination zwischen den untersuchten Individuen. Diese Art der Diskriminationsleistung des Test ist aber nicht auf die Umsetzung einer Metrisierung einer dispositionellen Eigenschaft in ein Meßinstrument zurückzuführen, sondern nur das Resultat der rechnerischen Synthese der numerisch symbolisierten Antworten auf die unterschiedlich schwierigen Testitems. Deswegen läßt sich ein konkreter "Messwert" in einem Test nur in Bezug auf seine formale bzw. rechnerische Synthese rekonstruieren und nicht in Bezug auf seine empirische Konstitution als Produkt der Auseinandersetzung des untersuchten Individuums mit der Prüfungssituation.

Das Vorgehen, die Funktionswertbestimmung von Tests über die durch sie erzeugte Varianz in der untersuchten Population festzulegen, steht außerdem im Gegensatz zu einigen syntaktischen Grundannahmen der klassischen Testtheorie, die für die Objektivierung des schulischen Beurteilungsprozesses herangezogen wird (FISCHER, 1974, S. 26-32). Diese besonderen, modellimmanenten Probleme bei der Anwendung formal-methodischer Prinzipien in der Beurteilung von Schülern können hier nicht diskutiert werden. Auf sie hinzuweisen ist aber zumindest deshalb wichtig, weil sie auch aus der Prioritierung des Diskriminationsprinzips erwachsen und sich in der inadäquaten Anwendung methodischer Regeln zur Testkonstruktion und Testbewertung die Verselbständigung der Entsubjektivierungsbestrebungen manifestiert und die Interpretationsbeliebigkeit der Anwendungsergebnisse konkretisiert.

Durch die Prioritierung des Diskriminationsprinzips wird die für die Bestimmung von dispositionellen Schülereigenschaften konstitutive Bedingung, daß die als gleich zu klassifizierenden Verhaltensprodukte von Schülern in Prüfungssituationen Ausdruck von invarianten Transformationsprozessen im Schüler sein sollen, auf einen formalen Aspekt reduziert: ein bestimmter Ausprägungsgrad einer dispositionellen Schülereigenschaft - und damit die Disposition selbst auch - wird dann als existent angenommen, wenn sein durch die Klassifikation der Verhaltensprodukte bestimmtes Verhältnis zu den anderen Schülern in einer Zweidurchführung des gleichen Tests oder in einer Durchführung eines ähnlichen Tests wiederholt auftritt, d.h., wenn seine durch Test festgelegte Position innerhalb der untersuchten Population über mehrere Untersuchungen hinweg invariant bleibt. Damit wird die empirische Begründung einer dispositionellen Eigenschaft durch den Aufweis einer statistischen Korrelation zwischen zwei Reihen von Testwerten ersetzt, die sich aus der wiederholten Anwendung des gleichen Verfahren ergeben. Zugleich dient dieses Ergebnis der Bestätigung der Behauptung, daß die Methode der Beurteilung der Schüler, der Test, zuverlässig arbeite.

Unabhängig von den Schwierigkeiten, die in der praktizierten Verallgemeinerung der so erhaltenen Beurteilungsergebnisse stecken, ist zu fragen, wie man in der empiristisch orientierten Diagnostik eine empirische Bedeutung solcher Urteile zu legitimieren versucht. In Bezug auf die praktizierte Diagnostik in der Schule stellt sich die Frage im Kontext der Selektionsentscheidungen, für die die Beurteilungsergebnisse herangezogen werden sollen. Deshalb muß für den Diagnostiker der Nachweis der empirischen Bedeutung von Urteilen über Schüler im Nachweis ihrer prognostischen Relevanz liegen und somit in der Bestätigung der Behauptung, daß dispositionelle Eigenschaften nicht nur über den Zeitraum invariant sind, der zwischen der Erst- und Zweitudurchführung des gleichen Tests liegt, sondern über den Zeitraum der Selektionsalternative. Mit der Überprüfung der prognostischen Relevanz einer dispositionellen Eigenschaft wird damit zugleich die inhaltliche Dimensioniertheit des als reliabel ausgewiesenen Tests aufzuzeigen versucht. Zum Nachweis der prognostischen Relevanz von Tests bedient man sich in der Testtheorie und Testkonstruktion des Verfahrens der Bestimmung der prognostischen Validität. Es unterscheidet sich in formaler Hinsicht von der Reliabilitätsbestimmung dadurch, daß an die Stelle der Zweitudurchführung des gleichen Tests oder der Durchführung eines äquivalenten Tests ein Instrument oder Beurteilungsvorgang für die Erfassung der dispositionellen Eigenschaft treten, welche in einer inhaltlichen Beziehung zu den in der dazwischenliegenden Phase stattgefundenen Ausbildungsprozessen stehen sollen. Von den Ausbildungsprozessen wird zwar angenommen, daß sie zur Entfaltung und somit Veränderung der dispositionellen Eigenschaften beitragen; die Methode zur Überprüfung der prognostischen Relevanz des reliablen Tests erlaubt aber nur die Erfassung eines invarianten Aspektes der unterstellten dispositionellen Eigenschaft. Dies hängt damit zusammen, daß auch für den zweiten Test oder Beurteilungsvorgang, die am Ende der Ausbildungsphase zur Anwendung kommen, die gleichen Probleme für die empirische Inter-

pretation der Beurteilungsergebnisse gelten wie für den Eingangstest, dessen prognostische Gültigkeit untersucht werden soll. Auch er kann die untersuchten Individuen lediglich in das abstrakte Verhältnis von Lösern und Nichtlösern zueinander setzen, das sich bei seiner wiederholten Anwendung im besten Fall als invariant erweisen kann. Ergibt sich im Verfahren der prognostischen Validitätsbestimmung des Eingangstests durch Korrelation zwischen seinen Ergebnissen und denen des Endtests bzw. Kriteriumstest ein relativ hoher numerischer Zusammenhang, so besagt das nur, daß die abstrakten Beziehungen zwischen den Individuen der untersuchten Population über den Zeitraum der Ausbildungsphase hinweg invariant geblieben sind bzw., daß die Ausbildungsprozesse in der dazwischenliegenden Phase zur Veränderung dieser Verhältnisse zwischen den Individuen nichts beigetragen haben. Wenn man bei der Interpretation solcher Beurteilungsergebnisse von der zusätzlichen Schwierigkeit absieht, daß nur die in die Ausbildungsalternative zugelassenen Schüler die Invarianzerscheinung konkretisieren können, dann kann die Interpretation der korrelativen Beziehung zwischen Eingangstest und Kriteriumstest nur aus der Spezifität der dazwischenliegenden Ausbildungsphase bezogen werden. Die Korrelation selbst drückt nur die Invarianz der formalen Beziehungen zwischen den untersuchten Schülern aus.

In diesem Dilemma äußert sich die fehlende Bestimmung der inhaltlichen Validität von objektivierten Beurteilungsverfahren. Da es nicht möglich ist, die Beurteilungsverfahren auf empirisch begründbare Eigenschaftsdimensionen zurückzuführen, muß ihre prognostische Relevanz durch die methodische Rekonstruktion der Gemeinsamkeiten von mehreren aufeinanderfolgenden Beurteilungsprozessen eingeholt werden - in der methodischen Rekonstruktion invarianter formaler Beziehungen zwischen den untersuchten Individuen. Der Versuch der inhaltlichen Aufklärung der Invarianz durch empiristisch orientierte Untersuchungen der diagnostischen Prozesse und der Unterrichtsprozesse, die zur Stabilisierung der Beziehungen zwischen den Schülern beitragen, mündet selbst

in die Konzipierung von komplizierten formalen Modellen, deren Interpretation die Frage nach der grundsätzlichen Entsubjektivierbarkeit des Beurteilungsprozesses aufwirft.

WESTMEYER (1972) versucht in seiner "Logik der Diagnostik" ein Modell zu entwickeln, mit dem die verschiedenen Determinanten der prognostischen Validität syntetisiert werden sollen. Die Konsequenzen aus seinem Modell deuten auf die dazu kontrastierende Simplizität in der derzeitig noch immer praktizierten Objektivierung von Schülerbeurteilungsprozessen und machen nachträglich die Beliebigkeit verständlich, mit der Testergebnisse und Zensuren interpretiert werden. In den gängigen Versuchen zur Objektivierung der Schülerbeurteilung beschränkt man sich auf das Auffinden einer funktionalen Beziehung zwischen einem Prädiktor und einem Kriterium. Bei der Betrachtung dieser Versuche aus der Perspektive des Ansatzes von WESTMEYER wird die Frage wichtig, von welchen Bedingungen in der Diagnostik bei der Verwendung der Validität als funktionale Beziehung zwischen Prädiktor und Kriterium abstrahiert wird bzw. über welche Bedingungen generalisiert wird. WESTMEYERs Ansatz ist für eine empiristisch orientierte Analyse diagnostischer Fragestellungen deshalb von besonderer Bedeutung, weil darin eine Reihe von Prinzipien und Forderungen empiristischer Wissenschaftstheorie ernst zu nehmen versucht werden und deshalb die Interpretationsschwierigkeiten in Bezug auf diagnostische Aussagen präzise hervortreten.

Nimmt man eines der wichtigsten Ergebnisse der Analyse vorweg, so drückt sich eine stringente Verwissenschaftlichung der Diagnostik in der von HÖRMANN (1964) und von WESTMEYER geteilten Konsequenz aus, daß die Interpretation von Beurteilungsdaten, die man mit "empirisch" validierten Instrumenten erhält, nur bereichsspezifisch zulässig ist. Daraus wird noch einmal ersichtlich, daß die Forderung nach objektiver und objektivierter Schülerbeurteilung nicht unabhängig von der Möglichkeit einzulösen ist, die empirische Bedeutung der Beurteilungsdaten zu bestimmen, d.h. von der Möglichkeit, die bereichsspezifische

Gültigkeit standardisierter Beurteilungsverfahren anzugeben. Entscheidend ist nicht, daß verschiedene Beurteiler zu gleichen Beurteilungsdaten gelangen, sondern in der inhaltlichen Interpretation der Beurteilungsdaten übereinstimmen.

Im Gegensatz dazu hat man sich bisher in der klassischen Testtheorie damit zufrieden gegeben, Objektivität, Reliabilität und Validität von Beurteilungsverfahren getrennt voneinander durch Anwendung bestimmter Überprüfungsmethoden zu beschreiben und war deshalb letztlich nicht in der Lage, die Verwendbarkeit der Beurteilungsinstrumente in intersubjektiver Übereinstimmung zu begründen. In den neueren, testtheoretischen Ansätzen für die Schülerbeurteilung wird zwar dem Problem der bereichsspezifischen Gültigkeit von Beurteilungsverfahren der Form nach Rechnung getragen, indem man an die Stelle von normorientierten Tests kriterienorientierte Tests zu setzen versucht (REISCHMANN, 1975)¹⁴); in anbetracht der fehlenden Qualifikationsbestimmungen für die Ausbildung in den existierenden Selektionsalternativen wie Hauptschule, Realschule und Gymnasium dürfte es aber unmöglich sein, kriterienorientierte Beurteilungsverfahren für Selektionsentscheidungen zu optimieren.

Da das Problem der empirischen Gültigkeit von objektivierten Beurteilungen nur immanent im Rahmen der empiristischen Orientierung diagnostischer Forschung diskutiert werden soll, ist zu fragen, wie sich hier die Bereichsspezifität darstellt. Allgemein testtheoretisch gesehen, hat ein diagnostisches Instrument immer dann empirische Gültigkeit, wenn es in der Tat jene dispositionellen Eigenschaften mißt, die es zu messen vorgibt. Diese tautologische Bestimmung wird in der praktizierten Testkonstruktion dadurch aufzulösen versucht, daß man die Ergebnisse der Testanwendung mit einem Gültigkeits- bzw. Validitätskriterium korreliert und die Höhe des Korrelationskoeffizienten als Maß der empirischen Bedeutung der Testergebnisse

14) Vgl. die Konkretisierungsversuche für die Schulleistungsmessung bei PRICE (1972).

interpretiert. Deswegen gibt es so viele empirische Bedeutungen wie hinreichend hohe Korrelationen mit verschiedenen Kriterien gefunden werden. Setzt man z.B. voraus, daß zwischen den Erfolgskriterien der beiden Selektionsalternativen "Hauptschule" und "Gymnasium", die in der Regel in Noten ausgedrückt werden, ein qualitativer Unterschied besteht, dann hat ein mit beiden Kriterien korrelierender Übertrittstest zwei verschiedene empirische Bedeutungen, jedenfalls nach dem zugrundegelegten Modell. Im konkreten Fall einer Einzelbeurteilung könnte also nicht entschieden werden, ob ein Schüler für die Hauptschule oder für das Gymnasium geeignet ist.

In seiner Analyse des Modells der empirischen Validierung zeigt nun WESTMEYER, daß die Menge der empirischen Bedeutungen durch die Anzahl der unterschiedlichen Kriterien keinesfalls erschöpft ist. Wenn der Validitätsbestimmung die Schlußweise zugrundegelegt wird, daß vom Testverhalten auf das Verhalten außerhalb der Testsituation geschlossen wird und das Kriterium des erschlossenen Verhaltens futurisch ist, sich also erst nach einem bestimmten Zeitraum erfüllt, so ist nach WESTMEYER die Validität ein siebenstelliges Prädikat mit folgenden Variablen: t (Test), c (Kriterium), p (Population, für die die Validität gelten soll), u (Umgebungsbedingungen für die Gültigkeit), v (Versuchsleitertyp) und z (Zeitbereich der Gültigkeit).

Höhe und Bedeutung der Validität sind neben dem Kriterium abhängig von der Personenklasse, auf die das Verfahren angewendet wird, von den Umgebungsbedingungen und dem Versuchsleitertyp bei der Durchführung des Tests und vom Zeitbereich, für den die Validität invariant bleiben soll. Darüberhinaus hängt die Validität natürlich vom formalen und strukturellen Aufbau des Beurteilungsverfahrens ab, z.B. von der Anzahl der Items, ihrer Form, ihrer Schwierigkeit, Trennschärfe usw. Jede dieser einzelnen Komponenten kann mehrfach variieren und bedingt dadurch eine differentielle Validität des Beurteilungsverfahrens. WESTMEYER kalkuliert an einem Beispiel die Menge möglicher Validitäten von Tests. Für vier verschiedene Personenklassen,

fünf verschiedene Kriterien und sechs verschiedene Items beträgt die Anzahl möglicher Validitäten 2048 (ebd., S. 61). Das Modell ließe sich ohne Schwierigkeiten um einige Komponenten erweitern und entsprechend verkomplizieren. Darauf kommt es aber nicht an. Wesentlich ist vielmehr der Umstand, daß die verschiedenen bereichsspezifischen Validitäten von objektivierten Beurteilungsverfahren verschiedene empirische Bedeutungen der Beurteilungsdaten erpräsentieren und diese Bedeutungen von festgelegten oder von bliebig rekonstruierbaren Bedingungen anhängen.

In der gegenwärtig praktizierten Objektivierung der Schülerbeurteilung werden die aus diesen Modellannahmen für die Validitätsbestimmung entstehenden Problem schlicht übergangen. Beim vorherrschenden Validierungsvorgehen erfahren vier der sechs benannten Bedingungsvariablen eine ausschließlich methodische Problematisierung:

1. Auf die Optimierung des Prädiktors t (Test) konzentrieren sich die meisten Forschungsaktivitäten. Im Vordergrund steht die methodische Bewältigung von methodisch bedingten Schwierigkeiten bei der empirischen Validitätsbestimmung, insbesondere der prognostischen Validität. Grundsätzlich erscheint dabei die Validität nur als Funktion der beiden Variablen t (Test) und c (Kriterium). Alle anderen Variablen finden keine explizite *p a r a m e t r i s c h e* Berücksichtigung.
2. Das Kriterium wird in der Regel nur insoweit problematisiert, als es unzuverlässig bzw. nicht objektiv im testtheoretischen Sinne ist. Die Analyse der inhaltlichen Validität des Kriteriums begegnet erheblichen Schwierigkeiten, da sie die Definierbarkeit der Lern- und Qualifizierungsziele des jeweiligen Unterrichts voraussetzt. Die Versuche durch kriterienorientierte Messung eine Lösung der Probleme herbeizuführen, bleiben entweder unzulänglich oder erweise sich als Spezialfälle der normorientierten Messung (BÜSCHER, 1974).
3. Die Variable v (Versuchsleitertypus) wird dadurch zu kontrollieren versucht, indem man die Beurteilungsverfahren zugunsten eines unterstellten intersubjektivitätszuwachses des Urteils sukzessiv standardisiert und einer methodisch verselbständigten, testtheoretischen Objektivitätsprüfung unterzieht. Dabei gehen genau jene Beurteilungsinformationen verloren, die noch am ehesten die aktive Auseinandersetzung des Schülers mit der diagnostischen Anregungssi-

tuation beinhalten. Zudem widerspricht die standardisierte Diagnose im Prinzip allen psychologischen Untersuchungen über die Aufnahme, Verarbeitung und Wiedergabe von Informationen (vgl. NEISSER, 1967).

4. Schließlich soll unter der Forderung nach Vergleichbarkeit der Beurteilungen die Populationsspezifität der Messung methodisch eliminiert werden (vgl. RASCH, 1960).

Völlig unberücksichtigt bleiben im praktizierten Validierungsvorgehen die Variablen u und z , also die Einflüsse von situativen Bedingungen und ihren Veränderungen in der Zeit auf die Möglichkeit der Prognose eines zu quantifizierenden Lernerfolges. Das bedeutet, daß sie für die Bestimmung der Validität eines Beurteilungsverfahrens als irrelevant oder konstant betrachtet werden. Berücksichtigt man von u und z nur einen Teilaspekt, nämlich die Unterrichtsgestaltung und den Unterrichtsverlauf, sowie die beide Faktoren bestimmenden Bedingungen - Curriculum, Lehrmaterial, Lehrer, soziale Organisation der Klasse, frühere Differenzierungen usw. -, so werden beim praktizierten Validierungsvorgehen genau jene Einflußgrößen ausgeklammert, die Gegenstand organisierter Bildung und Ausbildung sind.

Das Vorgehen abstrahiert daher von allen möglichen Einflüssen - schulischen oder außerschulisch-schulischen - auf den Lernerfolg der Schüler. Es bestätigt sich somit in der Betrachtung des gängigen Validierungsverfahrens, was sich bereits in der Untersuchung der Objektivitätsabsicherung gezeigt hat: Beurteilungsdaten, die mit Hilfe objektivierter Verfahren gewonnen werden, sind ebensowenig **e i n d e u t i g** inhaltlich interpretierbar wie Zensuren. Da sich die Anzahl möglicher Validitäten eines standardisierten Beurteilungsinstrumentes in Abhängigkeit von den ausgewählten Kriterien, den Umgebungsbedingungen der Testdurchführung, der spezifischen Personenklasse, auf die der Test angewendet wird, dem Versuchsleiter und dem intendierten Zeitbereich der Validität beliebig vergrößern läßt, kann man über die inhaltliche Bedeutung der mit dem Beurteilungsinstrument aktivierten dispositionellen Schülereigenschaften inner-

halb des Modells keine Entscheidung treffen. Ergibt sich zwischen den Daten einer Eingangsbeurteilung für eine Selektionsalternative und den Lernerfolgsdaten am Ende der Ausbildung in der Selektionsalternative eine relativ hohe Korrelation und besitzt somit das Instrument für die Eingangsbeurteilung eine hinreichende, spezifische Validität, so kann man nur folgendes behaupten: Durch das Beurteilungsinstrument werden die *untersuchten* Schüler zueinander in ein Verhältnis gesetzt, das über den Zeitbereich der Eingangsbeurteilung und der Lernerfolgsbeurteilung invariant geblieben ist und dessen noch unbekannte Invarianzbedingungen (die verschiedenen Komponenten, die die Validität beeinflussen) zu untersuchen sind. Die empirische Bedeutung dieses Verhältnisses läßt sich nicht angeben.

Nun besitzen aber standardisierte Beurteilungsverfahren gegenüber der Zensurengebung den "Vorzug", daß der bei ihrer Anwendung ablaufende diagnostische Prozeß, soweit er durch das Verfahren festgelegt ist, wiederholt werden kann. Die Standardisiertheit des diagnostischen Prozesses als Voraussetzung für seine partiell kontrollierbare Wiederholung ist im übrigen die einzige objektive Bedingung für die Verringerung subjektivistischer Einflüsse auf die Beurteilung. Sie eröffnet die Chance, die noch unbekannten Invarianzbedingungen des beurteilten Verhältnisses zwischen den Schülern für den vorliegenden Zeitbereich der Validität zu untersuchen.

Das gilt allerdings nur mit einer sehr bedeutsamen Einschränkung: Die Komponente der Personenklasse, für die die Validität des Instrumentes festgestellt wurde, entzieht sich der experimentellen Kontrolle, d.h., einmal beurteilte und in einer bestimmten Selektionsalternative ausgebildete Schüler stehen für eine Wiederholung des diagnostischen Prozesses nicht mehr zur Verfügung. Das hat zur Konsequenz, daß nur die anderen Komponenten wie Kriterium, Umgebungsbedingungen usw. festzulegen sind und ihr Einfluß auf die Validität überprüft werden kann. Bei der wiederholten Anwendung eines standardisierten Beurtei-

lungsverfahrens bleiben daher genau jene Komponenten in ihrem Einfluß auf die Validität unberücksichtigt, die letztlich Gegenstand der gesamten Beurteilungsbemühungen ist, die Spezifität der Personenklasse bzw. die Gruppe der konkreten, untersuchten Schüler. Wenn sich also über verschiedene Personenklassen, z.B. verschiedene Gruppen von Gymnasiasten, hinweg ein relativ hoher korrelativer Zusammenhang zwischen Eingangsbeurteilung und Lernerfolgsbeurteilung zeigt, dann kann dieser Effekt nur durch die zu untersuchenden anderen Komponenten rekonstruiert werden. Ihn durch die Konstanz von dispositionellen Eigenschaften erklären zu wollen wäre zirkulär. Er ist vielmehr nur durch die Gleichheit der anderen Komponenten für alle untersuchten Personenklassen zu erklären.

Die dem Validierungsmodell von WESTMEYER eigenen Annahmen erzwingen daher eine gegenüber der konventionelle Interpretation veränderte Funktionsbestimmung objektivierter Beurteilungsverfahren. Sie kann hier nur als konsequente Verwendungsmöglichkeit von solchen Verfahren angedeutet werden. Mit standardisierten Beurteilungsverfahren läßt sich überprüfen, ob extern festgelegte Beurteilungsbedingungen auf den diagnostischen Prozeß unabhängig von verschiedenen Personenklassen oder Schülergruppen und einzelnen untersuchten Individuen langfristig den gleichen Einfluß haben. Soweit die festgelegten Beurteilungsbedingungen (Kriterium, Umgebung, Versuchsleiter, Zeitbereich) in ihrer Auswirkung rekonstruierbar und verallgemeinerbar sind, kann man von der Validität eines Beurteilungsinstrumentes sprechen. Ferner können objektivierte Verfahren in einem eingeschränkten Sinn als Instrumente zur Beurteilung des Unterrichts fungieren, wenn die Lernerfolgskriterien thematisch bzw. stofflich auf die Ausbildung in der betreffenden Selektionsalternative bezogen sind. Da solche Lernerfolgskriterien die einzige Grundlage für eine unterrichtsinhaltliche Interpretation der Beurteilungsdaten darstellen, drückt ein relativ hoher korrelativer Zusammenhang zwischen Eingangsbeurteilung und Lernerfolgsbeurteilung die Tatsache aus, daß der Unterricht in der Ausbil-

dungsalternative zur Aufrechterhaltung eines diskriminativen Verhältnisses zwischen den Schülern beiträgt.

Die Fortschreibung von Eingangsunterschieden zwischen den Schülern durch Ausbildungsmaßnahmen setzt voraus, daß in der Ausbildung nichtindividualisierte, also für alle Schüler standardisierte Unterrichtsformen zur Anwendung kommen. Ein valides Beurteilungsinstrument für die Selektion bildet somit Verhältnisse zwischen den Schülern ab, die durch den Unterricht zeitlich verlängert werden. Solche Unterrichtsstandards zu rekonstruieren, könnte eine künftige Aufgabe einer empiristischen Unterrichtsforschung sein (vgl. HURRELMANN, 1971). Andernfalls ist die Validierung von Beurteilungsinstrumenten als wesentlicher Teilprozeß gegenwärtiger Verwissenschaftlichung der Schülerbeurteilung nicht mehr als eine Formalisierung des tradierten und intersubjektiv inkonsistenten Beurteilerverhaltens in der Schule.

3. Schlußfolgerungen

3.1 Ergebnisse der Versuche zur Entsubjektivierung der Schülerbeurteilung

Der empiristische Ansatz von WESTMEYER zur empirischen Deutung von Urteilen über Schüler in Form der prognostischen Validitätsbestimmung von objektivierten Beurteilungsprozessen ist das Resultat einer konsequenten, intendierten Entsubjektivierung des diagnostischen Prozesses. Er steht im Grad seiner logischen Ableitung und methodologischen Begründung am Ende einer fast schon zur Tradition gewordenen Tendenz zur Verwissenschaftlichung interpersonaler Beurteilung. Er repräsentiert im Bereich der Diagnostik die paradigmatische Bezogenheit empiristisch orientierter Sozialwissenschaft auf die Forderung nach Entsubjektivierung wissenschaftlicher Aussagen, wie sie sich in den exakten Naturwissenschaften in der geschichtlichen Entwicklung durchsetzte und in der empiristischen Wissenschaftstheorie systematisch reflektiert wurde.

Es wurde im Verlauf der Arbeit des öfteren darauf verwiesen, daß die programmatische Funktion des Entsubjektivierungsparadigmas innerhalb der Entwicklungsgeschichte der exakten Naturwissenschaften erst zu Beginn des zwanzigsten Jahrhunderts wirksam wird und nur aus der besonderen Bezogenheit naturwissenschaftlicher Forschung auf ihren Gegenstandsbereich begriffen werden kann. In Bezug auf die Verwissenschaftlichungstendenzen in der Diagnostik in der Schule, die als ein Teilbereich der empiristischen Orientierung der Sozialwissenschaften verstanden werden müssen, wurde die Gegenstandsbezogenheit der Forschung in einem entsprechenden Sinne bestritten und durch den Aufweis und die Erklärung der Interpretationsschwierigkeiten zu widerlegen versucht, die im intersubjektiv inkonsistenten Gebrauch von Urteilen über Schüler in Erscheinung treten und im Verständnis von Intersubjektivität selbst.

Die auftretenden Interpretationsschwierigkeiten haben ihren systematischen Ausgangspunkt in zwei Bedingungen:

1. Es besteht innerhalb der Forscher- und Diagnostikergemeinschaft kein intersubjektiver Konsens im Gebrauch der Forderung nach intersubjektiver Konsistenz im Gebrauch wissenschaftlicher Aussagen bzw. in der Forderung nach "Objektivität" solcher Aussagen.
2. Die Wissenschaftler- und Diagnostikergemeinschaft, soweit sie empiristisch orientiert ist oder sich zu einer Form empiristischer Forschung bekennt, bemüht sich kaum um eine systematische Reflexion jener Forschungsregeln und Wissenschaftsprinzipien, zu der sie ihre Orientierung verpflichten müßte.

Daraus resultiert zum einen, daß alle Reflexionstätigkeiten als "empirische Forschung" verstanden werden, die sich als Gegensatz zu Spekulation und subjektivistischen Einfallsreichtum definieren und nicht ohne Verwendung existierender, formalisierter Methoden und formalisierter methodologischer Prinzipien auskommen zu können glauben; zum zweiten resultiert daraus der Versuch empiristisch orientierter Wissenschaftler, verfügbare Methodensysteme auf beliebige Gegenstandsbereiche anzuwenden, ohne die Adäquanz solcher Methoden aus einer inhaltlichen Gegenstandsanalyse begründet zu haben. Äußere Erscheinung dieses umgekehrten Verhältnisses von Gegenstandsanalyse und methodischer Verallgemeinerung ist die Trennung von Syntax und Semantik in der diagnostischen Methodologie und die von methodologischer Reflexion und angewandter diagnostischer Forschung (FISCHER, 1974).

Die Funktion methodischer Regeln für die wissenschaftliche Tätigkeit und von methodologischen Prinzipien für die Legitimation dieser Tätigkeiten wurde mit dem Begriff "Forschungsparadigma" gekennzeichnet, wie ihn KUHN in seiner "Struktur wissenschaftlicher Revolutionen" entwickelte und wie ihn POPPER, LAKATOS und SCHEFFLER in seiner Bedeutung für die Wissenschafts-

entwicklung interpretierten. Diese Art der Interpretation scheint im Gegensatz zur Rezeption des Paradigmbegriffes durch STEGMÜLLER in seiner Wirksamkeit in den empiristisch orientierten Sozialwissenschaften gültig zu sein. Die strikte Orientierung eines Teils der Forschergemeinschaft der Diagnostiker und Beurteiler im Bereich der Verwissenschaftlichung der Schülerbeurteilung an einzelnen Regeln empiristisch wissenschaftlicher Tätigkeit dient hier der Objektivierung des Forschungs- und Beurteilungsprozesses um jeden Preis, um den kritisierten Subjektivismus des Lehrerurteils zu eliminieren. Die wissenschaftlichen Regeln zur Entsubjektivierung bzw. Objektivierung der Beurteilung von Schülern haben in der Zwischenzeit die Funktion einer "normativen Methodologie" erhalten, die erfolgreiches, wissenschaftliches Arbeiten in Aussicht stellt und den "normalen" Wissenschaftler und Diagnostiker dazu verpflichtet, seine Fragestellungen darunter zu subsumieren und die Ergebnisse seiner Reflexionsbemühungen dadurch zu legitimieren. Das Paradigma der Entsubjektivierung hat dadurch der formalisierenden empiristischen Forschung im Bereich der Diagnostik in der Schule eine gewisse, scheinbare Anschaulichkeit zurückgegeben, die in der Selbstverständlichkeit der Verwendung eingeübter methodischer Vorgehensweisen in der Forschung begründet liegt,

Die Herausarbeitung dieses psychologischen und sozialpsychologischen Faktors als eine Determinante der Wissenschaftsentwicklung, die in der Selbstverständlichkeit der Verwendung methodischer Regeln wissenschaftlicher Tätigkeit liegende Anschaulichkeit abstrakter Reflexion, was das wesentliche Anliegen von KUHN. In der Form der widersprüchlichen Anschaulichkeit, die in der Gewohnheit der Anwendung von reglementierten Denkweisen, Begriffen und Systemen liegt, mag KUHN letztlich den Grund gesehen haben, von einer Komponente "irrationalen" Verhaltens von Wissenschaftlern zu sprechen. BLJACHER & VOLYN-SKAJA (1974) nennen solche scheinanschaulichen Denkgewohnheiten "operationelle Strategeme" der Denktätigkeit. "Obwohl

hin und wieder Überlegungen dahingehend geäußert wurden, daß die Steigerung der Abstraktion des heutigen Wissens vom Verlust seiner Anschaulichkeit begleitet wird, so gibt es doch zuverlässige und in ausreichendem Maße überzeugende Beweise dafür, daß sogar im Wissen von hoher Abstraktionsstufe irgendeine 'anschauliche Ebene' vorhanden ist, die vor dem geistigen Auge des Forschers als implizite Voraussetzung der Theorie oder zumindest als Einschätzung ihrer Glaubwürdigkeit schwebt. ... Manchmal können Erscheinungen anschaulich werden, die von ihrer Substanz her keineswegs anschaulich sind. Den Status der Anschaulichkeit bekommen nicht nur Bilder, Vorstellungen, Modelle usw., sondern auch die operationellen Strategeme der Denktätigkeit" (ebd., S. 96f). Sofern solche Gewohnheiten das Reflexionsverhalten innerhalb einer Wissenschaftlergemeinschaft über einen bestimmten Zeitraum zu prägen beginnen und in Form von verallgemeinerten, institutionalisierten Regeln wissenschaftlicher Tätigkeit ihren Niederschlag finden, beeinflussen sie das Forscherverhalten in zwei Richtungen: die gegenständlich konstitutiven oder semantischen Grundlagen für ihre Entwicklung werden nicht mehr hinterfragt und ihre Anwendung und Ausdifferenzierung wird zum Gegenstand der Kommunikation unter den Wissenschaftlern.

Diese beiden Erscheinungsformen der Funktion von methodischen Regeln wissenschaftlicher Tätigkeit in ihrer Bedeutung für die Entsubjektivierung der Beurteilung von Schülern waren als Kennzeichen der Tendenz zur Verwissenschaftlichung der Diagnostik in der Schule zu untersuchen. Der besondere Bezug der Entsubjektivierungstendenzen in der Schülerbeurteilung zu bestimmten Forderungen aus der empiristischen Wissenschaftstheorie ergab sich dabei aus dem Anspruch der Wissenschaftler- und Diagnostikergemeinschaft, die Schülerbeurteilung auf den Boden empirischer Forschung zu stellen. Deshalb mußte untersucht werden, was diese Gemeinschaft unter dem Resultat der Entsubjektivierung, der "Objektivität" wissenschaftlicher Aussagen, im Hinblick auf empiristische Forderungen versteht und woraus

sich die Forderung nach Entsubjektivierung des Schülerbeurteilungsprozesses ableitet.

Ausgangspunkt der Untersuchung ist die Feststellung, daß innerhalb der empiristisch orientierten Forscher- und Diagnostikergemeinschaft Konsens in Bezug auf die Forderung nach "Objektivität" von Urteilen über Schüler besteht. Sie soll im Vollzug der Verwissenschaftlichung der Schülerbeurteilung hergestellt werden. Die systematische Betrachtung des Begriffes der Objektivität aus der Sicht empiristischer Wissenschaftstheorie zeigt, daß die Forscher- und Diagnostikergemeinschaft nur über einen formalen Begriff von Objektivität verfügt (s. 1.2). Dies äußert sich in der Existenz von Methoden zur Überprüfung der intersubjektiven Übereinstimmung von Beurteilungsergebnissen verschiedener Beurteiler bzw. zur Überprüfung, ob intersubjektive Übereinstimmung vorhanden ist oder nicht; sie läßt aber keine empirische Interpretation bzw. empiristische Rekonstruktion von Urteilen über Schüler zu und auch keine Erklärung des Phänomens der fehlenden intersubjektiven Übereinstimmung. Für die Gewinnung von objektiven Urteilen über Schüler wären diagnostische Prozesse durchzuführen, die auf dem übereinstimmenden Gebrauch einer Beobachtungssprache zur Klassifikation des Schülerverhaltens in diagnostischen Situationen aufbauen und in der intersubjektiv konsistenten Deutung von invarianten Verhaltensmustern, den dispositionellen Eigenschaften von Schülern als prognostisch relevante Lernvoraussetzungen, münden. Daß diese Forderung an die Objektivität von Urteilen in der praktizierten Diagnostik in der Schule nicht erfüllt wird, liegt nicht nur an der Verkürzung empiristische Theorienbildung und -bestätigung, sondern an dem Fehlen einer pädagogisch-psychologischen Entwicklungstheorie und ihrer gesellschaftlichen Begründung im Rahmen institutionalisierte Bildung und Ausbildung.

Trotz aller fehlenden empirischen Begründbarkeit von Beurteilungsdaten wird aber in der Verwissenschaftlichung der praktizierten Diagnostik in der Schule am Forschungsparadigma der

Entsubjektivierung festgehalten. Kennzeichnend für seine Wirksamkeit ist der fehlende Gegenstandsbezug in der Analyse des diagnostischen Prozesses und seine Verselbständigung in Form von institutionalisierten Regeln methodischen Vorgehens. Die sich darin ausdrückenden Gewohnheiten wissenschaftlicher Betrachtung des Gegenstandsbereiches schulischer Diagnostik werden in den Zusammenhang mit drei historischen Bedingungen für ihre Entstehung zu stellen versucht: der psychotechnisch-diagnostischen "Bewährungsprobe" des Entsubjektivierungsparadigmas in außerschulischen Bereichen, der Objektivierung der Kritik am Lehrerurteil durch das Entsubjektivierungsparadigma und der Funktion des Entsubjektivierungsparadigmas für die Technologisierung gesellschaftlicher Prozesse im allgemeinen (s. 1.3).

Die anschließende Auseinandersetzung mit dem KUHNschen Begriff des Paradigmas und der daran geübten Kritik verfolgt den Zweck, den Begriff zu präzisieren, seine allgemeine Funktion in der Entwicklung der Wissenschaften auszuweisen und seine besondere für die Verwissenschaftlichung der Diagnostik in der Schule als verselbständigtes und erfolgsverheißendes Prinzip für das Verhalten von Wissenschaftlern und Diagnostikern zu skizzieren. Es zeigt sich dabei, daß die "rationale Rekonstruktion" der KUHNschen Position durch SNEED und STEGMÜLLER nur dann gültig ist, wenn in dem betreffenden Forschungsbereich eine kontinuierliche, auf die empirische Gegenstandsanalyse bezogene Theorienentwicklung und -begründung stattgefunden hat. Nur dann erhalten Methodologie und Methodik im Forschungsprozeß ihre konkrete Bedeutung und bedürfen keiner besonderen Legitimation für die wissenschaftliche Analyse; und nur dann ist der Begriff des Paradigmas im Sinne einer mathematischen Struktur als Theoriekern zu interpretieren. Andernfalls gilt die KUHNsche Interpretation, wonach Paradigmen Regelprinzipien für das Forscherverhalten darstellen, deren rationale Steuerungsfunktion zu bezweifeln ist (s. 1.4).

Hier ergibt sich für die Betrachtung der Entsubjektivierungsbestrebungen in Bezug auf die Schülerbeurteilung eine spezifische Situation. Einerseits mangelt es an einer kontinuierlichen, empirischen Theorienentwicklung und gegenstandsbezogenen Erforschung des diagnostischen Prozesses, sodaß die dennoch eingesetzten Methoden für die Rekonstruktion dispositioneller Eigenschaften in ihrer Gegenstandsadäquanz nicht empirisch begründet sind. Auf der anderen Seite wird aber ihre Relevanz für die Gewinnung von objektiven Aussagen über Schüler-eigenschaften als gegeben vorausgesetzt und durch den impliziten Anspruch, nach Prinzipien der empiristischen Wissenschaftstheorie vorzugehen, legitimiert. Für die Abschätzung der Relevanz der Regeln wissenschaftlichen Arbeitens im Gegenstandsbereich der Diagnostik in der Schule bleibt daher nur der Weg, die aus der empiristischen Wissenschaftstheorie beanspruchten wissenschaftlichen Prinzipien differenziert zu betrachten und zu fragen, ob die formale Subsumption diagnostischer Fragestellungen darunter zulässig ist. Aus dieser Vorgehensweise resultieren zwei, in Bezug auf die Verwissenschaftlichungstendenzen zu überprüfende Behauptungen:

1. In der methodischen Objektivierung der Schülerbeurteilung kommen formale Strukturen zur Anwendung, für die keine definierten Gegenstandsbereiche existieren, und
2. wenn dem so ist, dann können sich die Forscher und Diagnostiker intersubjektiv konsistent nur über die methodische Konstitution der Objektivierung unterhalten, nicht aber über deren empirische Bedeutung.

Beide Behauptungen sind voneinander abhängig. Die zweite kann dann belegt werden, wenn sich zeigen läßt, daß die erste zutrifft und man gleichzeitig die Tatsache berücksichtigt, daß die existierenden Verwissenschaftlichungsbestrebungen (s. 1.1) eine potentielle intersubjektive Verständigung zwischen Diagnostikern naheulegen scheinen.

Im Kontext selektionsfunktionaler Entscheidungen über die "Eignung" von Schüler für die institutionalisierten, weiterführenden Ausbildungszweige besteht für die Verwissenschaftlichung der Diagnostik in der Schule die zentrale Aufgabe in der objektivierten Rekonstruktion von dispositionellen Eigenschaften, die sich langfristig als prognostisch relevant erweisen sollen. Die Überprüfung der aufgestellten Behauptungen mußte deshalb an zwei Problemen erfolgen: an der praktizierten und versuchten Rekonstruktion dispositioneller Schülereigenschaften und somit an der Handhabung des Dispositionsbegriffes in der Diagnostik, und an den Versuchen zur Bestätigung der prognostischen Relevanz solcher Eigenschaften.

Der Dispositionsbegriff und seine Einführungsmöglichkeiten in die Wissenschaftssprache wurden in der empiristischen Wissenschaftstheorie ausführlich und differenziert diskutiert. Die wesentliche Frage kreist um die Funktionsbestimmung von Dispositionsbegriffen in der Theorie und um ihre empirische Deutung durch eine Beobachtungssprache. Ein zentrales Ergebnis der Diskussion ist, daß Dispositionsbegriffe auch dann nur partiell empirisch gedeutet werden können, wenn sich die Wissenschaftlergemeinschaft in einer intersubjektiv konsistent gebrauchten Beobachtungssprache verständigen kann. Das unter dieser einschränkenden Voraussetzung reflektierte Vorgehen zur Einführung von Dispositionsbegriffen wird auf diagnostische Fragestellungen zu übertragen versucht. Dagegen sprechen aber mehrere Gründe:

1. Der Anwendungsbereich für ein zu entwickelndes Modell für die Rekonstruktion dispositioneller Eigenschaften in der Diagnostik unterscheidet sich von dem der Naturwissenschaften durch die wechselseitige Beziehung zwischen zwei personalen Interaktionssystemen, dem des untersuchten Individuums und dem des Forschers. Der exakte Naturwissenschaftler steht nicht vor der Aufgabe, Regelmäßigkeiten in einem vom ihm getrennten personalen Interaktionssystem zu rekonstruieren; die objektiv empirische Grundlage seiner

Forschungstätigkeit sind vielmehr gesetzesartige Abläufe von Prozessen in der Natur. In die diagnostische Rekonstruktion von dispositionellen Schülereigenschaften sind deshalb die intendierten und aktiven Verarbeitungen von diagnostischen Situationen durch das untersuchte Individuum miteinzubeziehen; sie läßt sich nicht auf die Feststellung von Kontingenzen zwischen Klassen von Reizlagen und Klassen von Verhaltensprodukten in Prüfungssituationen beschränken. SCHMIDT (1974) fordert daher für die Rekonstruktion dispositioneller Eigenschaften einen Untersuchungsansatz, der die wechselseitige Beziehung beider Interaktionssysteme aufgreift. Für das Ergebnis der Untersuchung, die rekonstruierte dispositionelle Eigenschaft, soll deshalb gelten, daß es die Kongruenz der Verarbeitung von Reizlage und Verhaltensresultat durch das Interaktionssystem des untersuchten Individuums und das des Forschers zum Ausdruck bringt (s. 2.2).

Bleibt das Geschehen im Interaktionssystem des untersuchten Individuums unberücksichtigt - was sich in einer exemplarischen Veranschaulichung praktizierter, empiristisch orientierter Rekonstruktionsversuche zeigt -, so hat die Festlegung des Anwendungsbereiches pädagogischer Diagnostik eine tautologische bzw. widersprüchliche Struktur. Sie wird in das Belieben der Forscher- oder Diagnostikergemeinschaft gestellt, sodaß diese nur über die Form der Festlegung intersubjektiven Konsens herstellen kann (s. 2.2.1). Allgemein ist bereits aus der Perspektive der SCHMIDTschen Strukturierung der Rekonstruktion dispositioneller Eigenschaften die Übertragung des Verfahrens zur Einführung von Dispositionsbegriffen in die Wissenschaftssprache kontraindiziert.

2. Eine zweite grundsätzliche Schwierigkeit in der Rekonstruktion von dispositionellen Eigenschaften nach empiristischem Vorbild ergibt sich aus dem Umstand, daß Dispositionsbegriffe nur partiell empirisch gedeutet werden können und deshalb ihre Erkenntnisfunktionalität theoretisch abgesichert werden

muß. Während man hier in den exakten Naturwissenschaften auf bewährte Theorien und feststehende Strukturkerne von Theorien zurückgreifen kann und deshalb der empirisch ungedeutete Bedeutungsüberschuß von Dispositionsbegriffen pragmatisch hinreichend eingeschränkt werden kann, fehlt in der praktizierten Diagnostik in der Schule ein derartig gesicherter, theoretischer Hintergrund. Dies äußert sich in der intersubjektiv inkonsistenten Deutung von Beurteilungskategorien und in den erfolglosen Versuchen ihrer methodischen Objektivierung. Eine notwendige gesellschaftliche Begründung von Qualifikationen und ihrer Beurteilung muß durch subjektivistische, normative Präferenzen ersetzt werden. Auch hier kann sich die Forscher- und Diagnostikergemeinschaft nur über das Vorgehen in der Festlegung und Bevorzugung von Beurteilungskategorien einigen, nicht aber über ihre inhaltliche Begründung (s. 2.3.1).

Da das Fehlen der inhaltlichen Begründbarkeit von Beurteilungskategorien die Hauptursache für das Scheitern auch von nichtempiristischen Ansätzen zur theoretischen Fundierung von Schülerqualifikationen darstellt (vgl. MEYER, 1972), wird verständlich, warum sich die diagnostische Rekonstruktion von dispositionellen Eigenschaften in der verfahrenstechnischen Festlegung und formalen Synthese von subjektivistisch beeinflussten Beurteilungsvorgängen erschöpft. Am Beispiel der idealtypischen Rekonstruktion einer Aufsatzbewertung wird gezeigt, daß dort die formale Synthese dispositioneller Eigenschaften am deutlichsten in Erscheinung tritt und wegen ihrer relativ geringen methodischen Objektivierung der geeignetste Gegenstand der "Subjektivismuskritik" sein kann (s. 2.3.2).

3. Die idealtypische Rekonstruktion der konventionellen Zensurengebung am Beispiel der Aufsatzbewertung repräsentiert allerdings am besten die Vernachlässigung des Geschehens im Interaktionssystem des untersuchten Individuums bei der Rekonstruktion dispositioneller Eigenschaften. Sie reduziert

den Beurteilungsvorgang auf eine theoretisch unbegründete und empirisch ungedeutete Klassifikation von Verhaltensprodukten der Schüler in Prüfungssituationen und auf die rechnerische Synthese der numerisch symbolisierten Klassifikationen. Trotz größerer, methodischer Komplexität in der Festlegung und formalen Synthese von Beurteilungskategorien, geht aber die methodische Objektivierung der Schülerbeurteilung in Form von standardisierten Beurteilungsverfahren von der gleichen Voraussetzung bei ihren Modellbildungen aus. Dies wird in der Folge näher untersucht.

Bleibt das Geschehen im Interaktionssystem des untersuchten Individuums in der diagnostischen Rekonstruktion dispositioneller Eigenschaften unberücksichtigt, so stoßen die Versuche der methodischen Objektivierung der Schülerbeurteilung auf verschiedene Grenzen in Bezug auf die Einführung von Dispositionsbegriffen in Entsprechung zum empiristischen Vorgehen und in Bezug auf die Bestätigung der empirischen Gültigkeit von diagnostischen Aussagen.

Die diagnostischen Dispositionsprädikate erweisen sich im Gegensatz zu denen in der exakten Naturwissenschaft als dreistellige. Sie können nicht ohne Vorbestimmung der Individuen eingeführt werden, denen die dispositionelle Eigenschaft zukommen soll. Mit dem praktizierten Vorgehen in der Diagnosik durch die Konstruktion und Anwendung von standardisierten Beurteilungsverfahren, die auf einer Folge von Aussagen wie "Wenn eine bestimmte Aufgabe zur Lösung vorgelegt wird und sie wird gelöst, dann indiziert die die Existenz einer dispositionellen Eigenschaft" bestehen, wird zugleich die Vorbestimmung der Individuen und die Rekonstruktion der dispositionellen Eigenschaft vorgenommen. Das dreistellige Dispositionsprädikat wird auf ein zweistelliges verkürzt.

Aus der Unbestimmtheit der Individuen resultieren für die Möglichkeit einer beobachtungssprachlichen Fundierung dispositioneller Eigenschaften veränderte Bedingungen. In den

exakten Naturwissenschaften sind die erlebnismäßig aufweisbaren oder instrumentell-experimentell bedingten Veränderungen von Objekten in festgelegten Situationen die empirisch objektive Grundlage für die Einführung von Dispositionsbegriffen. In der Diagnostik entziehen sich die Objektveränderungen, die im Schüler ablaufenden Verarbeitungsprozesse, einer derartigen Beobachtung. Objektive Grundlage für die Rekonstruktbildung kann hier nur die diagnostische Situation und das Ergebnis ihrer Verarbeitung durch den Schüler, das Verhaltensprodukt sein. Werden zwischen den als gleich klassifizierten diagnostischen Situationen und Verhaltensprodukten systematische Beziehungen festgestellt, so fehlt für deren Interpretation als Ausdruck der Wirksamkeit invarianter, dispositioneller Eigenschaften die empirische Grundlage; der erlebnismäßig aufweisbare oder experimentell realisierbare, konstitutive Verarbeitungsprozeß, der durch eine intersubjektiv verständliche Beobachtungssprache verallgemeinert werden müßte.

Es besteht daher die Möglichkeit, daß die rekonstruierten Eigenschaften nichts weiter als reine, gedankliche Gebilde sind. Sie können deswegen entstehen, weil der Gemeinschaft der Forscher und Diagnostiker die empirisch objektive Grundlage für die Festlegung eines Anwendungsbereiches ihres Rekonstruktionsmodelles fehlt. Der konstatierbare intersubjektive Konsens kann sich deshalb auf die Festlegung und Anwendung des Rekonstruktionsmodelles als bloße Form beschränken (s. 2.4.1 und 2.4.2).

Auch der Versuch, die feststellbaren Regelmäßigkeiten zwischen der Darbietung von Reizklassen (Prüfungsaufgaben) und den Verhaltensresultaten in diagnostischen Situationen ohne Rekurs auf die im untersuchten Individuum ablaufenden Verarbeitungsprozesse empiristisch zu rekonstruieren, scheitert an dem Fehlen eines Verfahrens für die empirisch zu begründende Klassifikation der verschiedenen Verhaltensprodukte und der sie provozierenden Prüfungsaufgaben (s. 2.4.4).

An die Stelle der empirischen Begründung von diagnostischen Aussagen muß deshalb die Festlegung und Setzung von objektiven Bedingungen für eine vergleichbare Schülerbeurteilung treten. Sie vollzieht sich durch die Standardisierung des diagnostischen Prozesses und durch die Entsubjektivierung des Deutungsproblems durch die Bestimmung der prognostischen Gültigkeit von objektivierten Urteilen über Schüler.

Durch den Verwendungskontext der Schülerbeurteilung, die Verteilung von Schülern auf weiterführende, niveaueverschiedene Ausbildungsalternativen, ist die Form der Urteile wesentlich vorbestimmt: die Urteile müssen maximal zwischen den Schülern diskriminieren und Ausprägungsgrade dispositioneller Eigenschaften ausdrücken. Dieser Verwendungskontext ist auch für die konventionelle Zensurenggebung konstitutiv. Die daraus resultierende Forderung an die Form der Schülerbeurteilungen, quantifizierbare Eigenschaften abzubilden, ist aber über den Weg der Standardisierung des Beurteilungsprozesses scheinbar leichter zu erfüllen. Hier werden die zu verrechnenden Klassifikationskategorien nicht erst nachträglich aus dem vorliegenden Prüfungsergebnis rekonstruiert, sondern vorher in einzelnen, nach dem Prinzip maximaler Diskrimination von Testaufgaben konkretisiert. Die so vorgenommene Entsubjektivierung der Beurteilung wird allerdings um den Preis der Verletzung von wesentlichen Regeln zur Einführung von metrischen Begriffen in die Wissenschaftssprache nach empiristischem Maßstab erkaufte. Sie läßt die empirische Nichtinterpretierbarkeit von scheinquantifizierten Urteilen über Schüler nochmals hervortreten in der praktizierten Gleichsetzung von Metrisierung und Messung. (s. 2.4.5.2 und 2.4.5.3).

Innerhalb der Bandbreite der Unentscheidbarkeit über den empirischen Bedeutungsgehalt von Schülerbeurteilungen findet die methodische Objektivierung der prognostischen Relevanz und damit die Entsubjektivierung der Interpretation ihren Platz. Wegen der Nichterfülltheit der empiristischen Regeln für die Einführung von metrischen Begriffen kann die Annahme der Exis-

tenz und Stabilität von dispositionellen Schülereigenschaften bzw. von Ausprägungsgraden solcher Eigenschaften nicht mehr empiristisch rekonstruiert werden; sie wird vielmehr legitimiert durch die Wiederholbarkeit formaler Unterschiede zwischen den Schülern, die in den standardisierten Verfahren zum Ausdruck kommen. Als zuverlässig und prognostisch gültig gelten nur solche Verfahren, die die gleiche Diskrimination zwischen den Schülern wiederholt anzeigen.

Damit ist der Anwendungsbereich objektivierter Schülerbeurteilung de facto nur durch ein formales Prinzip festgelegt: das Prinzip maximaler Diskrimination zwischen den Schülern. Darüber besteht in der Gemeinschaft der Diagnostiker und Forscher hinreichend eindeutiger Konsens und auch - von verschiedenen test-theoretischen Ansätzen abgesehen - in der methodischen Differenzierung und verfahrenstechnischen Realisierung dieses Prinzips. Es zeigt sich allerdings an der Betrachtung des Interpretationsmodelles von WESTMEYER (1972), daß die Formalisierung der Deutung solcher Schülerbeurteilungen zu subjektivistischer Beliebigkeit führt: es gibt soviele Bedeutungen von objektivierten oder entsubjektivierten Beurteilungsdaten, wie man formale Konstitutionsbedingungen für die Standardisierung des Beurteilungsprozesses anzunehmen bereits ist.

3.2 Zur realen Notwendigkeit der Entsubjektivierung der Schülerbeurteilung

Die in der praktizierten Diagnostik in der Schule durchgesetzte Reduktion des Anwendungsbereiches objektivierter Schülerbeurteilung auf das formale Prinzip maximaler Diskrimination ist keinesfalls nur das Resultat naiv-empiristischer Forschungspraxis. Als äußerliche Form der Wirksamkeit des Entsubjektivierungsparadigmas in der empiristisch orientierten, diagnostischen Forschung ist sie zugleich eine Notwendigkeit, die sich aus der Funktion, Organisation und Struktur des selektiven Bildungswesens ergibt.

Von daher ist es nicht erstaunlich, daß die verschiedenen Formen objektivierter Beurteilungen bestimmte "Vorzüge" der Zensurenggebung fortschreiben und deren optimale Realisierung im Verlauf der weiteren Verwissenschaftlichung der gesamten Ausbildung ermöglichen könnten. Um den "positiven" Stellenwert standardisierter Beurteilungen zu verdeutlichen, sollen die Vorzüge der Zensurenggebung, wie sie ZIELINSKI (1961) zusammenstellte, im Überblick betrachtet werden.

Wenn man die positiven Momente der Funktion von Zensuren zusammenfassen versucht und in ihrer Gesamtheit den von ZIELINSKI strukturierten negativen Konsequenzen gegenüberstellt, so läßt sich der Zusammenhang der Befürwortungen und jener der Ablehnungen substantiell bestimmen. Diese substantielle Bestimmung äußert sich darin, daß die Aufrechterhaltung des Notensystems in all jenen Argumenten befürwortet wird, die im Hintergrund ein abstraktes Interesse formulieren, das Interesse der Institution "Bildung und Ausbildung" bzw. das von gesellschaftlichen Instanzen, die für die Beibehaltung der gegenwärtigen Struktur des Bildungswesens eintreten. Demgegenüber beziehen sich die negativen Momente immer auf das Interesse von konkreten Personen, seien es die von der Beurteilung betroffenen Schüler oder Erziehungsberechtigten oder die den Beurteilungsprozeß vollziehenden Lehrer.

Darüber kann auch nicht die Tatsache hinwegtäuschen, daß die Institution "Bildung und Ausbildung" von konkreten Personen getragen wird. Denn deren Handlungsspielraum ist durch den abstrakten Zusammenhang aller Festlegungen und materiellen Grundlagen bestimmt, die das organisierte Lernen in seiner gegenwärtigen Erscheinungsform darstellen; dazu gehören die rechtlichen Grundlagen und gesetzlichen Bestimmungen zur Veranstaltung von Bildung und Ausbildung, die Ausbildungsvorschriften und Kompetenzfestlegungen für alle Personen, die mit dem Bildungswesen zu tun haben, die finanziellen und infrastrukturellen Bedingungen für den Ausbau und die materielle Konkretisierung des Schulsystems, die Zielbestimmung und curriculare Bestimmung für den Unterricht, verwaltungsorganisatorische Maßnahmen usw. Für diesen abstrakten Zusammenhang in seinen jeweils regional modifizierten Erscheinungsformen, in denen Schüler Qualifikationen erwerben sollen, muß die Beurteilung der Schüler funktional sein.

Berücksichtigt man nun - trotz aller existierenden Kritik an der Qualifizierungsleistung des Bildungssystems für private und öffentliche Abnehmer von Schülern und Studenten -, daß das Ausbildungssystem in

der Steuerung und Verteilung der Schüler auf sozioökonomisch abgestufte Bildungs- und Lebenschancen, also in der Differenzierung und Diskriminierung der Schüler, und

der linearen Organisation, hierarchischen Strukturierung und - von Randerscheinungen abgesehen - der irreversiblen Durchführung des Steuerungs- und Verteilungsprozesses

"erfolgreich" arbeitet, und berücksichtigt man zum zweiten, daß beide Aufgaben die allgemeine Funktion des gegenwärtigen Bildungswesens festlegen, dann drücken die befürwortenden Argumente gegenüber der Zensurengebung im Zusammenhang die Funktionalität der praktizierten Diagnostik in der Schule aus:

1. Das Notensystem bestätigt formal das Funktionieren institutionalisierter Erziehung, indem es die Annahme verallgemeinerbarer bzw. abstrakter Unterschiede zwischen den Schü-

lern rekonstruiert.

2. Es ermöglicht die rationell und intersubjektiv zuverlässig erscheinende Organisation der Schülerbeurteilung zum Zweck der Aufteilung der Schüler. Die damit durchgesetzte, organisatorische "Objektivität" der Beurteilung erschwert das Rückgängigmachen einer Selektionsentscheidung in außerordentlichem Maße.
3. Es entspricht seiner Form nach dem "Rationale" des Selektionsprinzips und erfüllt somit die Bedingung einer adäquaten Darstellung abstrakter Unterschiedlichkeit.
4. Es besitzt durch die Nichtrekonstruierbarkeit des in die Zensur eingehenden Beurteilungsprozesses eine besondere **l e g i t i m a t o r i s c h e** Dokumentationskraft. Diese sowie die dritte Eigenschaft der Zensuren erleichtern die Durchführung des zweiten Aufgabenkomplexes wesentlich.
5. Es verhindert aus demselben Grund die inhaltliche Problematisierung der Interpretierbarkeit der Zensuren durch die Beurteilten bzw. durch die Erziehungsberechtigten. Die Berechtigung einer Vergabe von Bildungschancen ist von daher letztlich nicht in Frage zu stellen.
6. Es garantiert die Aufrechterhaltung eines formalen Konkurrenzleistungsbewußtseins, vermittelt dadurch das Bewußtsein einer Quasinaturhaftigkeit der dokumentierten Leistungsunterschiede und trägt dadurch zum reibungsloseren Ablauf des organisierten Lernens bei.

Mit der Kennzeichnung der beiden Aufgaben, der Verteilung der Schüler auf sozioökonomisch abgestufte Bildungs- und Lebenschancen und der linearen Organisation, der hierarchischen Strukturierung sowie der irreversiblen Durchführung des Verteilungsprozesses, ist das abstrakte und allgemeinste Interesse der Institution "Bildung und Ausbildung" an der Qualifizierung von Schülern umrissen. Die "positiven" Funktionsmomente der Zensurumgebung haben sich diesem Interesse gegenüber als funktional

erwiesen. Es kann also behauptet werden, daß die Zensurengebung insofern den Bedingungen des organisierten Lernens entspricht, als in ihr die beiden Aufgaben als Funktionsgröße, die ich hier die Formbestimmungen institutionalisierter Ausbildung nenne, enthalten sind. Diese beiden Formbestimmungen äußern sich auf der Beurteilungsebene in dem Bestreben,

- a) eine optimale Differenzierung zwischen den Schülern zu gewährleisten und
- b) den Differenzierungsprozeß rationell zu organisieren und zuverlässig zu gestalten.

Unter der Voraussetzung, daß die funktionale Entsprechung gesichert sein muß, kann man das bereits erwähnte Zitat von HELLER (1974) als eine Umschreibung beider Bestrebungen begreifen:

"Von unterschiedlichen Aufgabenschwerpunkten und entsprechend modifiziert eingesetzten Beurteilungstechniken, etwa im dreigliedrigen versus Gesamtschulsystem, abgesehen kulminieren sämtliche Beurteilungs- und Beratungsansätze letztlich um die Differenzierungsproblematik. An dieser Frage und ihrer Lösung führt m.E. kein Weg vorbei, wenigstens ist bislang kein funktionierendes Schulsystem, das diese Problematik ohne Schaden ausklammern könnte, bekannt" (ebd., S. 20).

Beide Bestrebungen sind somit im bestehenden Ausbildungssystem die *conditio-sine-qua-non* und legen in letzter Instanz den Verwendungskontext jeder Art von Schülerbeurteilung, durch Zensurengebung, durch Wortgutachten, durch Diagnosebögen, durch Beratungsverfahren und durch *objektivierte*, psychotechnische Verfahren fest. Bei der Entwicklung von neuen Verfahren zur Schülerbeurteilung und bei Modifikationen der gegenwärtig praktizierten Diagnostik in der Schule haben also Vorbestimmungen den Vorrang, deren inhaltliche Auffüllung, z.B. mit intersubjektiv vermittelbaren Klassifikationsgesichtspunkten zur Beurteilung von vergegenständlichten Schülerverhaltensweisen bzw. mit Rekonstruktionen des empirischen Konstitutionsprozesses dieser Verhaltensweisen, ein nachgeordnetes Problem darstellt.

Auf die Austauschbarkeit der inhaltlichen Auffüllung von objektivierten Beurteilungsprozessen hinzuweisen, erscheint mir dann besonders wichtig, wenn die These berechtigt ist, daß die empiristisch orientierte Rekonstruktion dispositioneller Eigenschaften von Schülern erfolglos bleiben muß.

Ich möchte deshalb zum Schluß die Frage stellen und unbeantwortet lassen, in wessen Interesse sich der Aufwand für eine Entsubjektivierung des Beurteilungsprozesses im Bereich der Diagnostik in der Schule lohnt.

Literaturverzeichnis

- AEBLI, H.: Entwicklungspsychologische Kriterien für die Auswahl von Curriculuminhalten. In: FREY, K. (Hrsg.): Kriterien in der Curriculumkonstruktion. Weinheim: Beltz, 1970, S. 49-55.
- AMTHAUER, R.: Intelligenzstrukturtest. Göttingen: Hogrefe, 1953.
- AUTORENKOLLEKTIV AM INSTITUT FÜR SOZIOLOGIE DER FU BERLIN: Klassenlage und Bewußtseinsformen technisch-wissenschaftlicher Lohnarbeiter. Zur Diskussion über die "Technische Intelligenz". Frankfurt: Europäische Verlagsanstalt, 1973.
- BAETHGE, M.: Ausbildung und Herrschaft. Unternehmerinteressen in der Bildungspolitik. Frankfurt: Europäische Verlagsanstalt, 1970.
- BAHRDT, H.P.: Industriebürokratie - Versuch einer Soziologie des industrialisierten Bürobetriebes und seiner Angestellten. Stuttgart, 1958.
- BECHERT, J., CLEMENT, D., THÜMMEL, W. & WAGNER, K.H.: Einführung in die generative Transformationsgrammatik. München: Hueber, 1971.
- BECKER, E. & JUNGBLUT, G.: Strategien der Bildungsproduktion. Frankfurt: Suhrkamp, 1972.
- BILDUNGSBERICHT '70: Bericht der Bundesregierung zur Bildungspolitik. Bonn: Heger, 1970.
- BLJACHER, E.D. & VOLYNKAJA, L.M.: Die Verallgemeinerung des physikalischen Weltbildes als historisches Moment des Erkenntnisprozesses. In: MARXISMUS DIGEST 18, Heft 2, 1974, S. 94-108.
- BLOOM, B.S. (ed): Taxonomy of Educational Objectives, Handbook I: Cognitive Domain. New York: David McKay, 1956.
- BOCK, R.D. & WOOD, R.: Test Theory. In: Annual Review of Psychology, 1971, 22, pp. 193-224.

- BÖHME, G., DAELE, W.v.d. & KROHN, W.: Die Finalisierung der Wissenschaft. In: DIEDERICH, W. (Hrsg.): Beiträge zur diachronischen Wissenschaftstheorie. Frankfurt: Suhrkamp, 1974, S. 276-311.
- BOESCH, E.E.: Die diagnostische Systematisierung. In: HEISS, R. (Hrsg.): Psychologische Diagnostik. Handbuch der Psychologie, 6. Band. Göttingen: Hogrefe, 1964, S. 901-929.
- BRIDGMAN, P.W.: The Nature of Some of Our Physical Concepts. In: British Journal for the Philosophy of Science, 1951, 2, pp. 257-272; 1951, 2, pp. 25-44; 1951, 2, pp. 142-160.
- BRINKMANN, H., BRÜCKNER, P., KRAHL, H.-J. & LAUERMANN, M.: Sinnlichkeit und Abstraktion. Prolegomena zu einer materialistischen Empirie. Wiesbaden: Focus, 1973.
- BRUNER, J.: Processes of Cognitive Growth. Worcester, Mass. 1968.
- BÜSCHER, P.: Einige testtheoretische Aspekte kriterienbezogener Leistungsmessung. In: HELLER, K. (Hrsg.): Leistungsbeurteilung in der Schule. Heidelberg: Quelle & Meyer, 1974, S. 137-157.
- COOMBS, P.H.: The World Educational Crisis - A System Analysis. Paris, 1967 (UNESCO).
- DAS ARGUMENT 88: 16. Jg., 1974, Heft 10-12.
- DEUTSCHER STÄDTETAG (Hrsg.): Die Zukunft der deutschen Städte. Berlin, 1925.
- DICK, F.: Kritik der bürgerlichen Sozialwissenschaften. Theorie und Praxis, Forschungsprozeß und Wissenschaftstheorie. Heidelberg: Sendler, 1974.
- DIEDERICH, W. (Hrsg.): Beiträge zur diachronischen Wissenschaftstheorie. Frankfurt: Suhrkamp, 1974.
- DOHSE, W.: Das Schulzeugnis - Sein Wesen und seine Problematik. Reihe: Pädagogische Studien - Bd. 10. Weinheim: Beltz, 1967.
- EDWARDS, W.: Behavioral Decision Theory. In: EDWARDS, W. & TVERSKY, A. (eds.): Decision Making. Middlesex: Harmondsworth, 1967, pp. 65-95.

- ENGELHARDT, M.v. & HOFFMANN, R.-W.: Wissenschaftlich-technische Intelligenz im Forschungsbetrieb. Frankfurt: Europäische Verlagsanstalt, 1974.
- EVERS, C.-H.: "Lernen tut zuweilen weh". In: betrifft:erziehung, 6. Jg., 1974, Heft 3, S. 45-47.
- FELLENIUS, V.: Berufs- und Arbeitspsychologie. In: KATZ, D. & KATZ, R. (Hrsg.): Handbuch der Psychologie. Basel/Stuttgart: Schwabe & Co., 1960, S. 573-587.
- FISCHER, G.: Einführung in die Theorie psychologischer Tests. Grundlagen und Anwendungen. Bern, Stuttgart, Wien: Huber, 1974.
- FLECHSIG, K.-H., GARLICH, A., HALLER, H.-D., HEIPKE, K. & SCHLÖSSER, H.: Probleme der Entscheidung über Lernziele. In: Programmirtes Lernen, Unterrichtstechnologie und Unterrichtsforschung, 7. Jg., 1970, S. 1-31.
- FODOR, J.A.: Could meaning be an r ? In: Journal of Verbal Learning and Verbal Behavior, 1965, Vol. 4, pp. 73-81.
- FOPPA, K.: Lernen Gedächtnis Verhalten. Ergebnisse und Probleme der Lernpsychologie. Köln, Berlin: Kiepenheuer & Witsch, 1965.
- FREUDENTHAL, H.: Lernzielfindung im Mathematikunterricht. In: Zeitschrift für Pädagogik, 20. Jg., 1974, Heft 5, S. 719-738.
- FRICKE, F.: Über Meßmodelle in der Schulleistungsdiagnostik. Düsseldorf: Schwann, 1972.
- FRIEDRICH, L. & KÖHLER, K. (Hrsg.): Zeugnisnoten und Numerus Clausus. Kronberg/Ts.: Scriptor, 1975.
- GAEDIKE, A.-K.: Determinanten der Schulleistung. In: HELLER, K. (Hrsg.): Leistungsbeurteilung in der Schule. Heidelberg: Quelle & Meyer, 1974, S. 46-93.
- GAGNE, R.M.: Die Bedingungen menschlichen Lernens. Hannover: Schroedel, 1969.
- GAUDE, P. & TESCHNER, W.-P.: Objektivierete Leistungsmessung in der Schule. Frankfurt: Diesterweg, 1970.
- GHISELLI, E.E. & BROWN, C.W.: Personal and Industrial Psychology. New York, Toronto, London: Mc Graw-Hill, 1955.

- GROFFMANN, K.J.: Die Entwicklung der Intelligenzmessung. In: HEISS, R. (Hrsg.): Psychologische Diagnostik. Handbuch der Psychologie, 6. Band. Göttingen: Hogrefe, 1964, S. 148-199.
- GUILFORD, J.P.: The Structure of Intellect. In: Psychological Bulletin, 1956, 54, pp. 267-293.
- GULLIKSEN, H.: Theory of Mental Tests. New York: Wiley, 1950.
- HALE, D.E.: Moderation. In: MacINTOSH, H.G.: Techniques and Problems of Assessment. London: Edward Arnold Ltd., 1974.
- HALLER, H.-D.: Fragen der Unterrichts- und Curriculumforschung - wissenschaftsgeschichtlich betrachtet. In: Zeitschrift für Pädagogik, 19. Jg., 1973, Heft 4, S. 572-582.
- HARTMANN, K., NYSSSEN, F. & WALDEYER, H. (Hrsg.): Schule und Staat im 18. und 19. Jahrhundert. Frankfurt: Suhrkamp, 1974.
- HARTOG, P. & RHODES, E.C.: An Examination of Examinations. London: MacMillan, 1936.
- HEISS, R.: Psychologische Diagnostik: Einführung und Überblick. In: HEISS, R. (Hrsg.): Psychologische Diagnostik. Handbuch der Psychologie, 6. Band. Göttingen: Hogrefe, 1964, S. 3-16.
- HELLER, K. (Hrsg.): Leistungsbeurteilung in der Schule. Heidelberg: Quelle & Meyer, 1974.
- HELMER, O.: Social Technology. New York: Basic Books, 1966.
- HÖRMANN, H.: Theoretische Grundlagen der projektiven Tests. In: HEISS, R. (Hrsg.): Psychologische Diagnostik. Handbuch der Psychologie, 6. Band. Göttingen: Hogrefe, 1964, S. 71-112.
- HÖRMANN, H.: Aussagemöglichkeiten psychologischer Diagnostik. Göttingen: Hogrefe, 1964.
- HOLZKAMP, Ch.: Statistik als ein Aspekt der wissenschaftlichen Methoden in der Unterrichtsforschung. In: INGENKAMP, K. (Hrsg.): Strategien der Unterrichtsforschung. Weinheim: Beltz, 1973, S. 61-97.
- HOLZKAMP, K.: Sinnliche Erkenntnis - Historischer Ursprung und gesellschaftliche Funktion der Wahrnehmung. Frankfurt: Fischer Athenäum, 1973.

- HOPF, D.: Forschungsstand, Forschungsschwerpunkte und Institutionalisierung der Pädagogischen Diagnostik. In: DEUTSCHER BILDUNGSRAT: Gutachten und Studien der Bildungskommission. Bildungsforschung, Probleme-Perspektiven-Proritäten, Band 51, Teil 2. Stuttgart: Klett, 1975, S. 217-263.
- HURRELMANN, K.: Unterrichtsorganisation und schulischer Organisationsprozeß. Weinheim: Beltz, 1971.
- INGENKAMP, K. (Hrsg.): Die Fragwürdigkeit der Zensurengebung. Weinheim: Beltz, 1971.
- INGENKAMP, K.: Pädagogische Diagnostik. Weinheim: Beltz, 1975.
- ITELSON, L.: Mathematische und kybernetische Methoden in der Pädagogik. Berlin/DDR: Verlag Volk und Wissen, 1967.
- KALVERAM, K.T.: Über Faktorenanalyse. Kritik eines theoretischen Konzepts und seine mathematische Neuformulierung. In: Archiv für Psychologie, Band 122, 1970, S. 92-118.
- KIESSLING, A.: Leistungsbeurteilung und Leistungsmessung. In: Zeitschrift für Pädagogische Psychologie, 1929, 30.
- KLEINMUNTZ, B. (ed.): Formal Representation of Human Judgment. New York: Wiley, 1968.
- KMK: Beschluß der ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik vom 3.10.68: Erläuterung der Notenstufen bei Schulzeugnissen und Einzelergebnissen in staatlichen Prüfungszeugnissen, Absatz II. In: Kultusministerkonferenz (Öffentl. A.) Erg.-Lfg. 12 v. 21.4.69, Nr. 671, S. 1.
- KÖTTER, L. & GRAU, U.: Zur Bedingtheit der uneinheitlichen Benotung von Schüleraufsätzen. In: Zeitschrift für experimentelle und angewandte Psychologie, 1965, 12, S. 278-301.
- KROPP, R.P. & STOKER, H.W.: The Construction and Validation of Tests of the Cognitive Processes as Described in the Taxonomy of Educational Objectives. Florida State University, 1966.
- KUDER, G.F. & RICHARDSON, M.W.: The Theory of the Estimation of Test Reliability. In: Psychometrika, 1937, 2, S. 151-160.
- KUHN, T.: Die Struktur wissenschaftlicher Revolutionen. Frankfurt: Suhrkamp, 1973.

- KUKLA, F.: Entwicklungsaspekte des Messens in der Psychologie. In: HIEBSCH, H. & SPRUNG, L. (Hrsg.): Aufgaben, Perspektiven und methodologische Grundlagen der marxistischen Psychologie in der DDR. Berlin/DDR: VEB Deutscher Verlag der Wissenschaften, 1973, S. 115-122.
- KUTSCHERA, F.v.: Wissenschaftstheorie I. Grundzüge der allgemeinen Methodologie der empirischen Wissenschaften. München: Fink, 1972.
- LAKATOS, I.: Falsification and the Methodology of Scientific Research Programmes. In: LAKATOS, I. & MUSGRAVE, A. (eds.): Criticism and the Growth of Knowledge. Cambridge: University Press, 1970, pp. 91-195.
- LANGFELDT, H.-P.: Die klassische Testtheorie als Grundlage standardisierter Schulleistungstests. In: HELLER, K. (Hrsg.): Leistungsbeurteilung in der Schule. Heidelberg: Quelle & Meyer, 1974, S. 96-136.
- LIENERT, G.A.: Testaufbau und Testanalyse. Weinheim: Beltz, 1961.
- LIETZMANN, W.: Über die Beurteilung der Leistung in der Schule. Mathematisches, Psychologisches, Pädagogisches. Leipzig-Berlin, 1927.
- MACKWORTH, J.F.: Vigilance and Attention. Harmondsworth: Penguin Books, 1970.
- MAGER, R.F.: Lernziele und Programmierter Unterricht. Weinheim: Beltz, 1971.
- MAGNUSSON, D.: Testtheorie. Wien: Deuticke, 1969.
- MESSNER, R.: Funktionen der Taxonomien für die Planung von Unterricht. In: Zeitschrift für Pädagogik, 16. Jg., 1970, Heft 6, S. 755-779.
- MEYER, H.L.: Einführung in die Curriculummethodologie. München: Kösel, 1972.
- MICHEL, L.: Allgemeine Grundlagen psychometrischer Tests. In: HEISS, R. (Hrsg.): Psychologische Diagnostik. Handbuch der Psychologie, 6. Band. Göttingen: Hogrefe, 1964, S. 19-70.
- MÖLLER, Ch.: Technik der Lernplanung. Weinheim: Beltz, 1969.
- NARSKI, S.I.: Dialektischer Widerspruch und Erkenntnislogik. Berlin/DDR: VEB Deutscher Verlag der Wissenschaften, 1973.

- NEISSER, U.: Cognitive Psychology. New York: Appleton-Century-Crofts, 1967.
- NEWELL, A. & SIMON, H.A.: An Example of Human Chess Play in the Light of Chess Playing Programs. In: WEINER, N. & SCHADE, J.P. (eds.): Progress in Biocybernetics, Vol. 2. Amsterdam: Elsevier, 1965.
- NEWELL, A., SHAW, J.C. & SIMON, H.A.: The Processes of Creative Thinking. In: GRUBER, H.E., TERRELL, G. & WERTHEIMER, M. (eds.): Contemporary Approaches to Creative Thinking. New York: Atherton Press, 1967, pp. 63-119.
- OFFE, C.: Bildungssystem, Beschäftigungssystem und Bildungspolitik - Ansätze zu einer gesamtgesellschaftlichen Funktionsbestimmung des Bildungssystems. In: DEUTSCHER BILDUNGSRAT: Gutachten und Studien der Bildungskommission. Bildungsforschung, Probleme-Perspektiven-Prioritäten, Band 50, Teil 1. Stuttgart: Klett, 1975, S. 215-252.
- OLDFIELD, R.C. & MARSHALL, J.C. (eds.): Language. Harmondsworth: Penguin Books, 1968.
- OSGOOD, C.E., SUCI, G.J. & TANNENBAUM, P.H.: The Measurement of Meaning. Urbana: University of Illinois Press, 1957.
- OTTEN, D.: Kapitalentwicklung und Qualifikationsentwicklung. Zum logischen und historischen Verhältnis von wirtschaftlich-technischer Entwicklung und Ausbildung des Arbeitsvermögens. Westberlin: Rossa, 1973.
- PLANCK, M.: Wissenschaftliche Autobiographie. Leipzig, 1928.
- POLLIO, H.R.: Associative Structure and Verbal Behavior. In: DIXON, T.R. & HORTON, D.L. (eds.): Verbal Behavior and General Behavior Theory. Englewood Cliffs, N.J.: Prentice Hall, 1968, pp. 37-66.
- POPPER, K.R.: Normal Science and Its Dangers. In: LAKATOS, I. & MUSGRAVE, A. (eds.): Criticism and the Growth of Knowledge. Cambridge: University Press, 1970, pp. 51-58.
- PREUSS, U.K.: Das Problem Bildung - Bildung und Qualifikation. In: Neue Sammlung, 1973, 1, S. 26-41.
- PROJEKTGRUPPE: Diagnostik in der Schule. München: Oldenbourg, 1973.
- PSYCHOLOGISCHE DIAGNOSTIK: HEISS, R. (Hrsg.): Psychologische Diagnostik. Handbuch der Psychologie, 6. Band. Göttingen: Hogrefe, 1964.

- RASCH, D.: Elementare Einführung in die mathematische Statistik. Berlin/DDR: VEB Deutscher Verlag der Wissenschaften, 1970.
- RASCH, G.: Probabilistic Models for Some Intelligence and Attainment Tests. The Danish Institute for Educational Research: Copenhagen, 1960.
- RASCH, G.: An Informal Report on a Theory of Objectivity in Comparisons. Proc. of the NUFFIC International Summer Session in Science at "Het Oude Hof", The Hague, 14.-28. Juli, 1966.
- RASCHERT, J.: Möglichkeiten und Grenzen der Forschung bei der Findung, Begründung und Auswahl von Lernzielen. In: DEUTSCHER BILDUNGSRAT: Gutachten und Studien der Bildungskommission. Bildungsforschung, Probleme- Perspektiven-Prioritäten, Band 51, Teil 2. Stuttgart: Klett, 1975, S. 55-85.
- REISCHMANN, J.: Der lehrgangsorientierte Test. In: Unterrichtswissenschaft, 1975, Heft 1, S. 50-60.
- RIEDLER, A.: Emil Rathenau und das Werden der Großwirtschaft. Berlin, 1916.
- RORSCHACH, H.: Psychodiagnostik. Bern: Huber, 1921.
- RUCH, G.M. & STODDARD, G.D.: Tests and Measurement in High School Instruction. New York: World Book Co., 1929.
- SCHEFFLER, I.: Science and Subjectivity. New York: Bobbs-Merrill, 1967.
- SCHMIDT, H.-D.: Zum Problem des Konstruktbegriffes in der empirischen Persönlichkeitsforschung und -diagnostik. In: BÖTTCHER, H.R., SEEGER, A. & WITZLACK, G. (Hrsg.): Psychodiagnostik - Probleme, Methoden, Ergebnisse. Berlin/DDR: VEB Deutscher Verlag der Wissenschaften, 1974, S. 20-24.
- SCHRÖTER, G.: Die ungerechte Aufsatzzensur. Bochum: Kamp, 1971.
- SCHULREFORM: Schulreform in Bayern: Band 1, Lehrpläne für die Grundschule, Orientierungsstufe und Hauptschule. München: Bayerisches Staatsministerium für Unterricht und Kultus, 1970.
- SCHWARZ, E.: Experimentelle und quasi-experimentelle Anordnungen in der Unterrichtsforschung. In: INGENKAMP, K. (Hrsg.): Strategien der Unterrichtsforschung. Weinheim: Beltz, 1973, S. 99-193.

- STARCH, D. & ELLIOT, D.C.: Reliability of Grading High School Work in Mathematics. In: School Review, Vol. 21, 1913, S. 254-259.
- STEGMÜLLER, W.: Probleme und Resultate der Wissenschaftstheorie und Analytischen Philosophie. Band II: Theorie und Erfahrung. Berlin: Springer; Studienausgabe Teil A, Teil B und Teil C, 1970; Teil D und Teil E, 1973.
- STÖHR, W.: Lehrerschaft und Arbeiterbewegung - Zur Organisationsgeschichte der gewerkschaftlichen Lehrerbewegung 1918-1933. In: DAS ARGUMENT 80, 1973, S. 170-207.
- STRITTMATTER, P. (Hrsg.): Lernzielorientierte Leistungsmessung. Weinheim: Beltz, 1973.
- SÜLLWOLD, F.: Schultests. In: HEISS, R. (Hrsg.): Psychologische Diagnostik. Handbuch der Psychologie, 6. Band. Göttingen: Hogrefe, 1964, S. 352-384.
- TACK, W.H.: Messung als Repräsentation empirischer Gegebenheiten. In: Zeitschrift für experimentelle und angewandte Psychologie, 1970, 17, S. 184-212.
- TENT, L.: Die Auslese von Schülern für weiterführende Schulen. Göttingen: Hogrefe, 1969.
- TOULMIN, S.: Voraussicht und Verstehen. Ein Versuch über die Ziele der Wissenschaft. Frankfurt: Suhrkamp, 1968.
- ULICH, D. & MERTENS, W.: Urteile über Schüler. Weinheim: Beltz, 1973.
- VAGT, R.: Forschungspraxis und Forschungsobjekt. In: HAAG, F., KRÜGER, H., SCHWÄRZEL, W. & WILDT, J. (Hrsg.): Aktionsforschung. Forschungsstrategien, Forschungsfelder und Forschungspläne. München: Juventa, 1972, S. 9-21.
- VUKOVICH, A.F.: Die Konstruktion psychologischer Tests. In: HEISS, R. (Hrsg.): Psychologische Diagnostik. Handbuch der Psychologie, 6. Band. Göttingen: Hogrefe, 1964, S. 113-144.
- WALLACH, M.A. & KOGAN, N.: Modes of Thinking in Young Children: A Study of the Creativity-Intelligence Distinction. New York: Holt, Rinehart & Winston, 1965.
- WATSON, R.I.: Historical Review of Objective Personality Testing: The Search for Objectivity. In: BASS, B.M. & BERG, I.A. (eds.): Objective Approaches to Personality Assessment. New York, 1959, S. 1-23.

WESTMEYER, H.: Logik der Diagnostik. Grundlagen einer normativen Diagnostik. Stuttgart: Kohlhammer, 1972.

ZIEGENSPECK, J.: Zensur und Zeugnis in der Schule. Hannover: Schrödel, 1973.

ZIELINSKI, J.: Macht und Ohnmacht der Zensuren. In: Pädagogische Rundschau, 15. Jg., 1961, Heft 1/2, S. 121-142.

Lebenslauf

Am 15. Juli 1943 wurde ich in Thorn (damals in Westpreußen) geboren. Gegen Ende des Krieges mußten meine Mutter, meine Schwester und ich nach Ilmendorf in die Nähe von Ingolstadt an der Donau übersiedeln. Mein Vater kehrte 1946 aus der Kriegsgefangenschaft heim und fand in München eine Beschäftigung in seinem Beruf als kaufmännischer Angestellter. Ich besuchte zunächst in Ilmendorf die Volksschule, mußte aber wegen des Umzugs nach München im Jahre 1951 die Schule wechseln. Den Besuch der Ludwigs-Oberrealschule in München schloß ich 1963 mit dem Abitur ab. Im Anschluß daran wurde ich zur Bundeswehr eingezogen und beendete meine zweijährige Dienstzeit als Leutnant der Reserve.

Von 1965 bis 1970 studierte ich an der Universität München Psychologie, Philosophie und Pädagogik. Das Diplom in Psychologie erhielt ich im Frühjahr 1970.

Unmittelbar danach arbeitete ich als wissenschaftlicher Assistent am Institut für Unterrichtsforschung der Pädagogischen Hochschule Augsburg auf dem Gebiet der kognitiven Psychologie. Ab 1.7.1971 war ich als wissenschaftlicher Angestellter im Projekt Schullaufbahnberatung am Institut für Pädagogik II der Universität München und am Institut für Unterrichtsforschung der Pädagogischen Hochschule Augsburg tätig. In dieser Zeit beschäftigte ich mich in erster Linie mit Fragen der pädagogischen Diagnostik.

Nach dem vertraglichen Ablauf des Forschungsprojektes Ende 1973 erhielt ich von der Universität München ein Graduiertenstipendium ab August 1974. Seitdem habe ich an der Abfassung der vorgelegten Dissertation gearbeitet und mich auf die Promotion vorbereitet.

Im Hinblick auf meine künftige berufliche Beschäftigung bin ich weiterhin an einer wissenschaftlichen Tätigkeit im Bereich der Bildungsforschung interessiert.

